

# Práctica 2 Competición Titanic en kaggle

Javier Maestre Deusto y Miguel López Marzabal

9 de junio, 2020

## Contents

<b>1. Descripción del dataset. ¿Por qué es importante y qué pregunta/problema pretende responder?</b>	<b>2</b>
<b>2. Integración y selección de los datos de interés a analizar.</b>	<b>2</b>
Trabajo con las diferentes variables . . . . .	2
Trabajo con Ticket . . . . .	2
Trabajo con Name . . . . .	5
Trabajo con Parch . . . . .	8
Trabajo con Embarqued . . . . .	8
Trabajo con Cabina . . . . .	9
Trabajo con Fare . . . . .	9
Trabajo con Edad . . . . .	13
Trabajo con Pclass . . . . .	22
<b>3. Limpieza de los datos.</b>	<b>23</b>
<b>3.1. ¿Los datos contienen ceros o elementos vacíos? ¿Cómo gestionarías cada uno de estos casos?</b>	<b>23</b>
<b>3.2. Identificación y tratamiento de valores extremos.</b>	<b>24</b>
<b>4. Análisis de los datos.</b>	<b>24</b>
<b>4.1. Selección de los grupos de datos que se quieren analizar/comparar (planificación de los análisis a aplicar).</b>	<b>24</b>
Análisis de las variables por métodos de filtrado . . . . .	24
Análisis de las variables por análisis de factores . . . . .	31
Análisis de las variables por Forward Selection, Backward Selection, Stepwise Selection . . . . .	33
Análisis de las variables por Backward Selection con paquete mlr . . . . .	33
Análisis de las variables por Recursive Feature Elimination Method (RFE) . . . . .	196
Análisis de las variables RandomForest . . . . .	197
<b>4.2. Comprobación de la normalidad y homogeneidad de la varianza.</b>	<b>199</b>
<b>4.3. Aplicación de pruebas estadísticas para comparar los grupos de datos. En función de los datos y el objetivo del estudio, aplicar pruebas de contraste de hipótesis, correlaciones, regresiones, etc. Aplicar al menos tres métodos de análisis diferentes.</b>	<b>199</b>
Predicción por Random Forest . . . . .	199
Predicción por KNN . . . . .	200
Predicción por regresión logística . . . . .	200
Predicción por redes neuronales y análisis de componentes principales . . . . .	217

## 1. Descripción del dataset. ¿Por qué es importante y qué pregunta/problema pretende responder?

Se trata de un dataset compuesto por los pasajeros del Titanic y cual fue su desenlace, si sobrevivieron o no. La finalidad del estudio es predecir cual fué el destino final de un conjunto de pasajeros en los cuales no hay datos del desenlace. Este estudio forma parte de una competición abierta en el portal de ciencia de datos Kaggle, por lo que está ampliamente documentado y existe una gran diversidad de métodos por los cuales se puede obtener el resultado. La competición está basada en el porcentaje de acierto de la predicción hecha. El dataset está compuesto por los siguientes campos:

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

**\*\* Variable Notes \*\***

pclass: A proxy for socio-economic status (SES) 1st = Upper 2nd = Middle 3rd = Lower

age: Age is fractional if less than 1. If the age is estimated, is it in the form of xx.5

sibsp: The dataset defines family relations in this way... Sibling = brother, sister, stepbrother, stepsister Spouse = husband, wife (mistresses and fiancés were ignored)

parch: The dataset defines family relations in this way... Parent = mother, father Child = daughter, son, stepdaughter, stepson Some children travelled only with a nanny, therefore parch=0 for them.

## 2. Integración y selección de los datos de interés a analizar.

En este apartado, a parte de la integración y selección de datos, vamos a realizar conversión(construcción de atributos, agregación?, normalización?, discretización?, jerarquía de conceptos?) y si es necesario reducción.

Unimos los conjuntos de train y test para analizar los datos al completo

### Trabajo con las diferentes variables

En este documento vamos a separar el trabajo hecho en cada variable, por tanto irán mezcladas cosas de preprocesado

#### Trabajo con Ticket

Partimos el campo ticket en prefijo y número de ticket. A partir del numero de ticket obtenemos grupos en función de la cantidad de numero del número de ticket y vemos como se relaciona con sobrevivir. Vemos

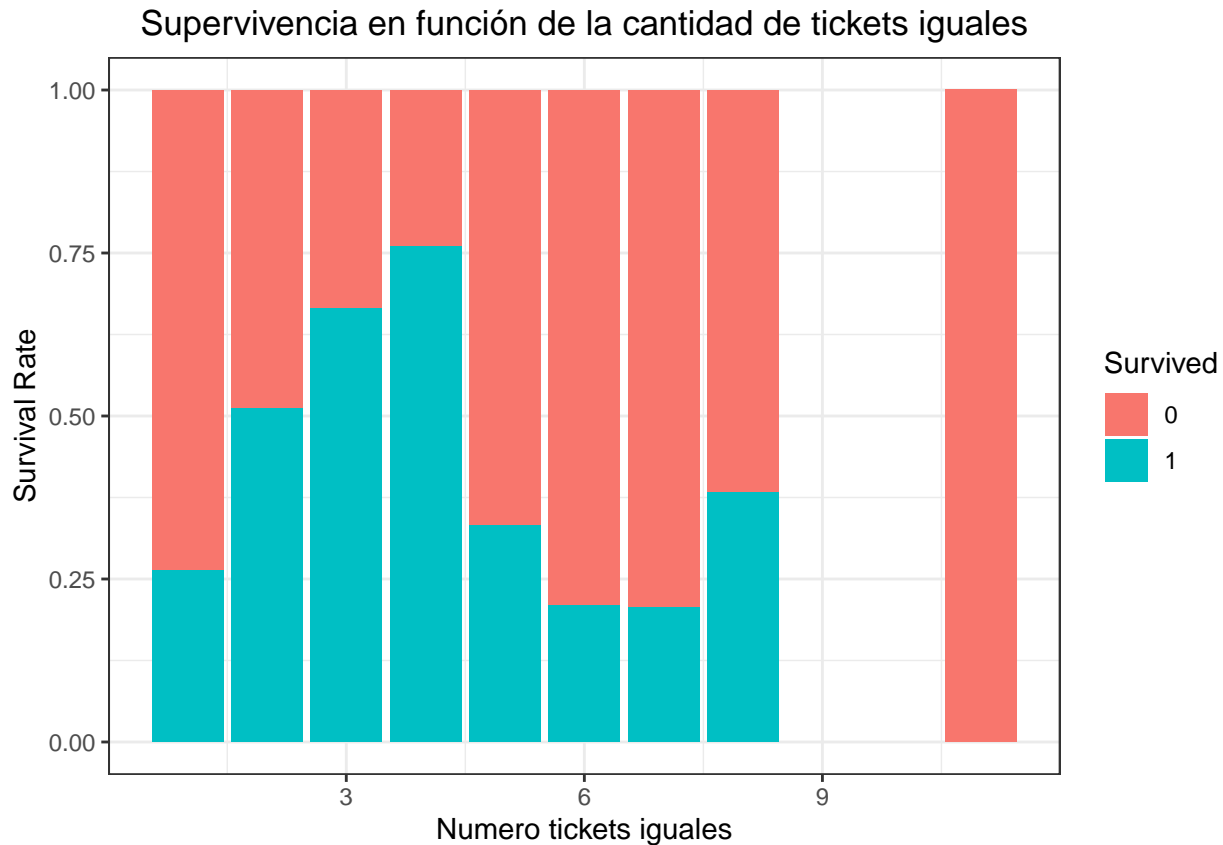
como se relacionan los prefijos del ticket con sobrevivir. Obtenemos una variable que nos indica la catidad de números de tickets(TicketFreq) iguales y vemos su relacion con sobrevivir. Hemos tenido que corregir 4 casos de valores perdidos en nuestra nueva variable debido a que estos casos no tenian número de ticket.

Stacked bar chart showing Survival Rate by Longitud numero ticket. The y-axis represents the Survival Rate from 0.00 to 1.00. The x-axis represents the Longitud numero ticket from 0 to 7. The legend indicates that red represents 'Survived' (0) and teal represents 'Not Survived' (1).

Longitud numero ticket	Survived (0)	Not Survived (1)
0	0.75	0.25
1	1.00	0.00
2	0.00	0.00
3	0.60	0.40
4	0.70	0.30
5	0.45	0.55
6	0.70	0.30
7	0.70	0.30

Stacked bar chart showing the survival rate by ticket prefix. The y-axis is 'Survival Rate' from 0.00 to 1.00. The x-axis is 'Prefijo Ticket' with categories: A4, A5, C, CA, COT, EN, F, FC, CP, PP, PS, S, SC, AH, BS, CP, AR, CS, GP, NO, NS, PD, NV, P, V, VE, P. The legend indicates 'Survived' with 0 (red) and 1 (teal).

Prefijo Ticket	Survived 0 (Red)	Survived 1 (Teal)
A4	1.00	0.00
A5	0.90	0.10
C	0.60	0.40
CA	0.65	0.35
COT	1.00	0.00
EN	1.00	0.00
F	1.00	0.00
FC	0.25	0.75
CP	0.40	0.60
PP	0.60	0.40
PS	0.50	0.50
S	1.00	0.00
SC	0.50	0.50
AH	1.00	0.00
BS	1.00	0.00
CP	0.55	0.45
AR	0.80	0.20
CS	1.00	0.00
GP	1.00	0.00
NO	1.00	0.00
NS	0.90	0.10
PD	1.00	0.00
NV	0.55	0.45
P	1.00	0.00
V	0.10	0.90
VE	0.70	0.30
P	1.00	0.00



### Trabajo con Name

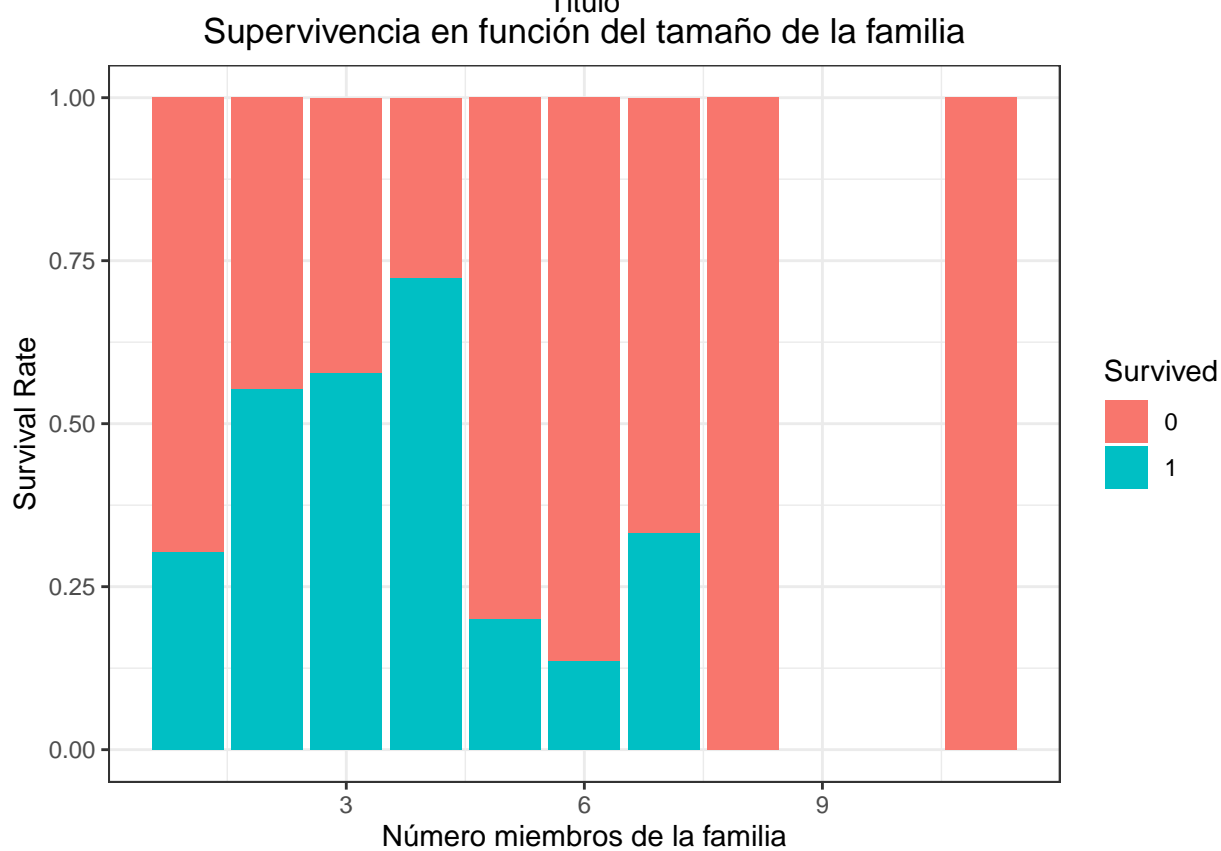
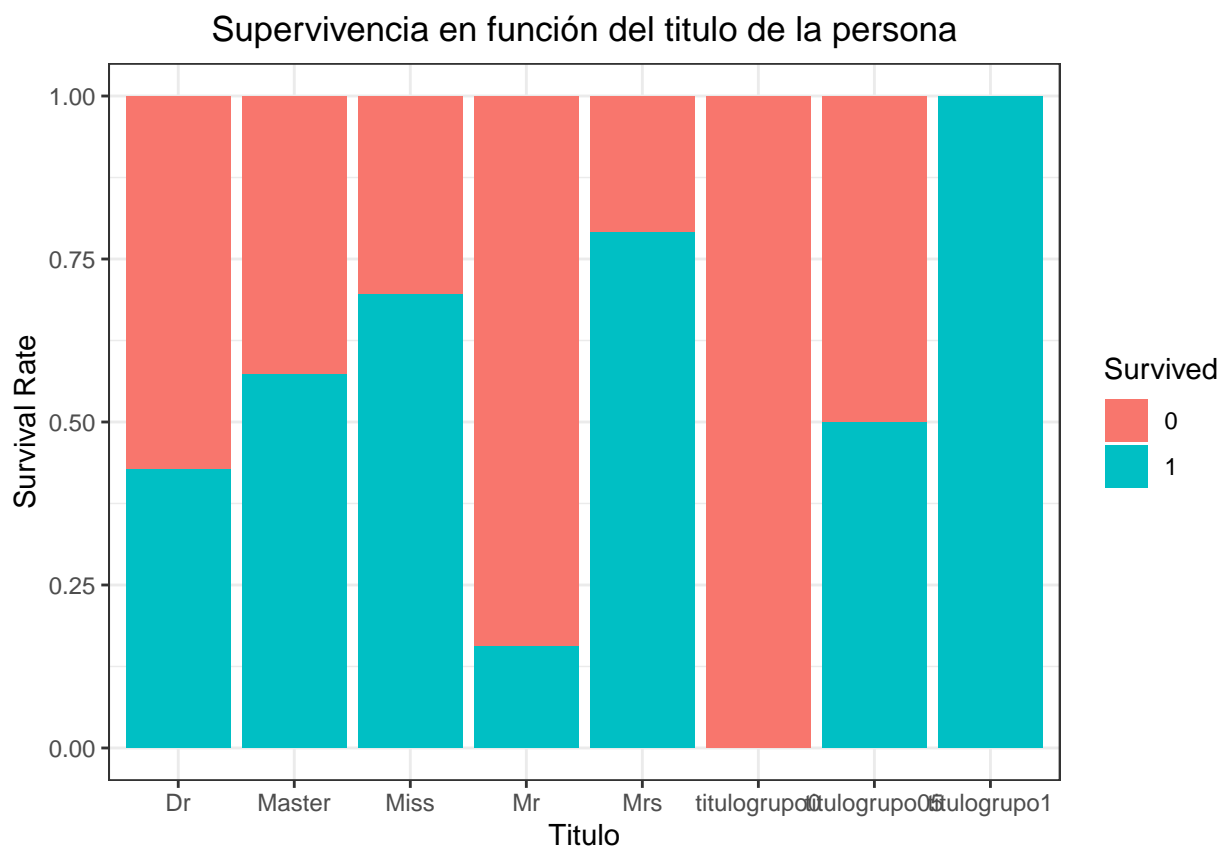
Partimos el campo name para obtener por un lado el Apellido del pasajero y por otro lado el título que recibe esa persona.

Agrupamos los títulos poco frecuentes en tres grupos según su probabilidad de supervivencia, los que todas las personas con ese título han muerto van a 'titulogrupo0', si han sobrevivido todos a 'titulogrupo1' y si la probabilidad es diferente al grupo 'titulogrupo05'. Comprovamos que la relación de los grupos con supervivencia.

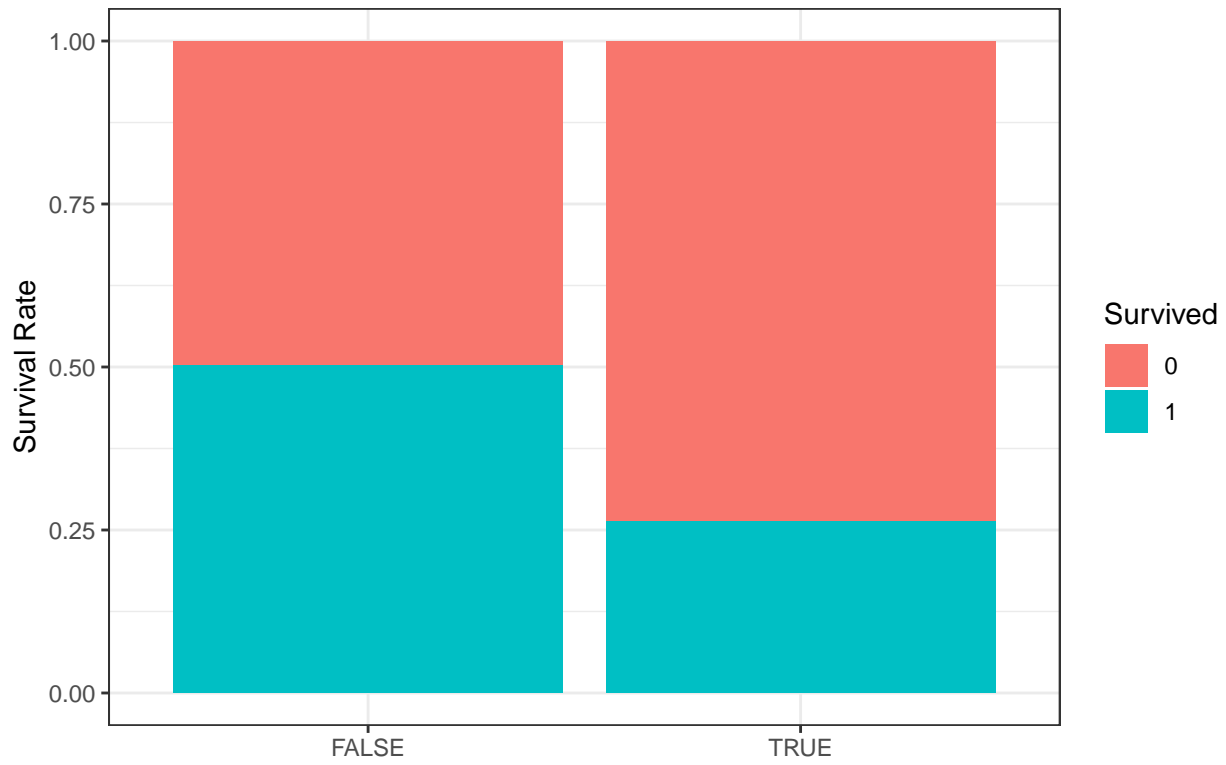
Creamos una nueva variable llamada FamSize que nos indica el número de familiares a bordo, esto se obtiene con 'SibSp + Parch + 1'. Mostramos gráficamente la relación con supervivencia.

Creamos una nueva variable 'LoneWolfs' que nos indica si una persona viaja totalmente sola, eso lo obtenemos con los elementos que cumplen esta condición 'FamSize == 1 & TicketFreq==1' y visualizamos la relación de la nueva variable con sobrevivir.

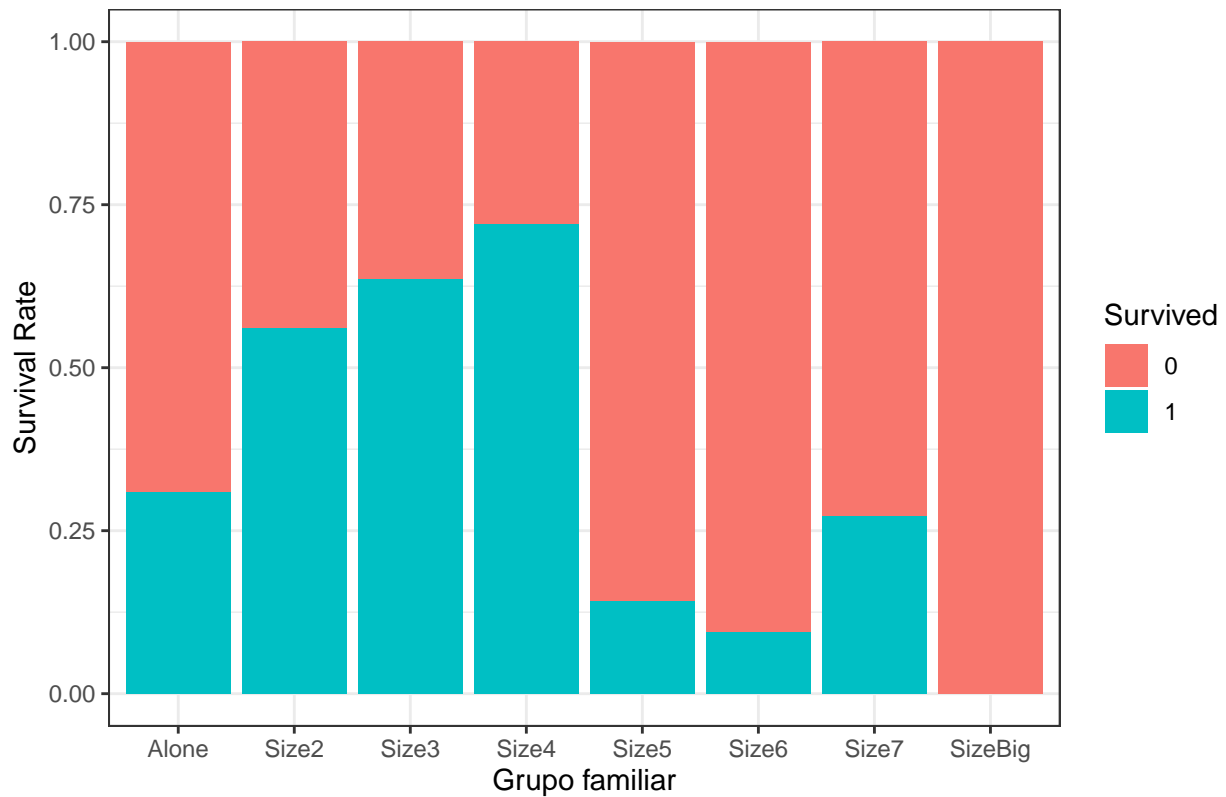
Vamos a agrupar las familias para ver exactamente cuantos familiares viajaban juntos. Esto lo haremos concatenando el Apellido con el número de familiares y el número de ticket. Hemos añadido el número de ticket porque si solo juntamos apellido y número de familiares se mezclaban familias que coincidían en apellido y numero de familiares pero viajaban con diferente ticket. Tras agruparlos correctamente creamos grupos genéricos según el número real de familiares que viajaban juntos. Finalmente visualizamos la relación de nuestra nueva variable 'FamilyIDTKGrouped' con la supervivencia.



Supervivencia en función de si viajan solos

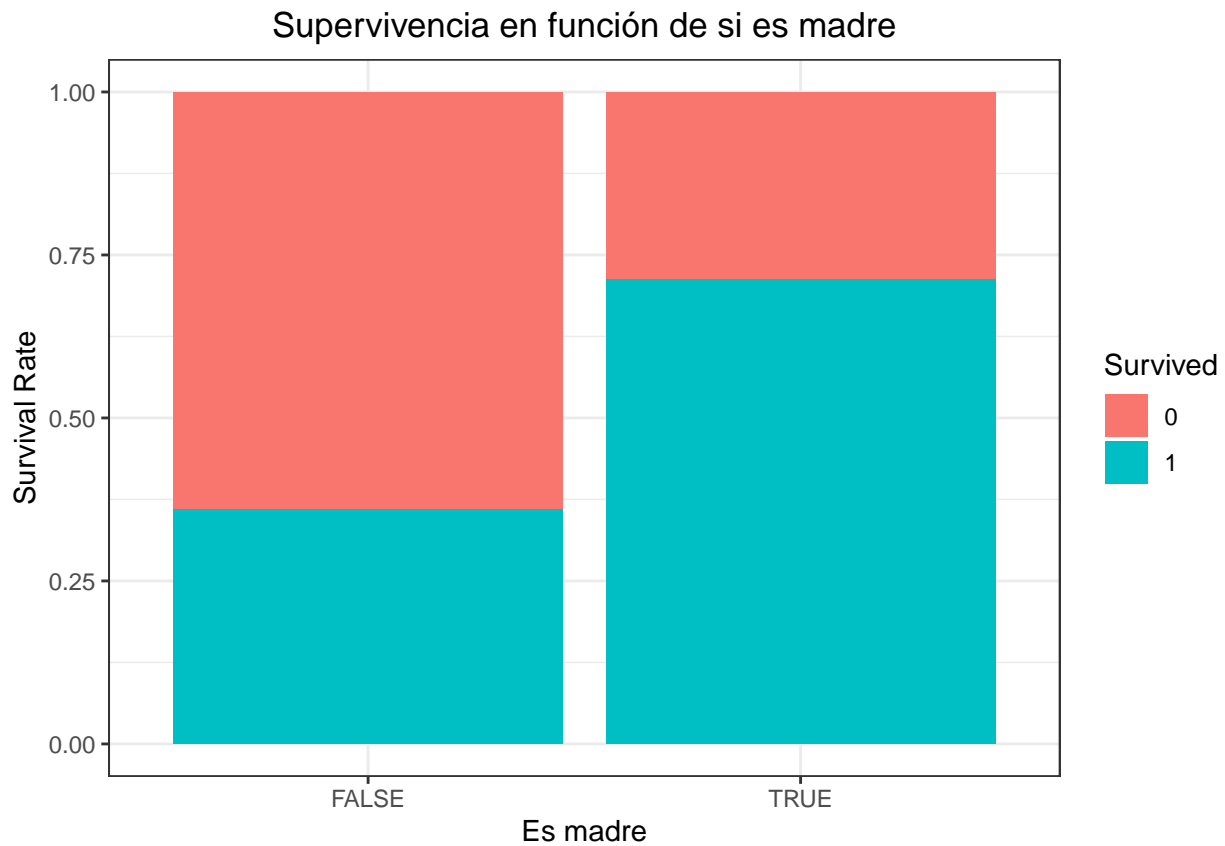


Viaja solo  
Supervivencia en función de su grupo familiar



### Trabajo con Parch

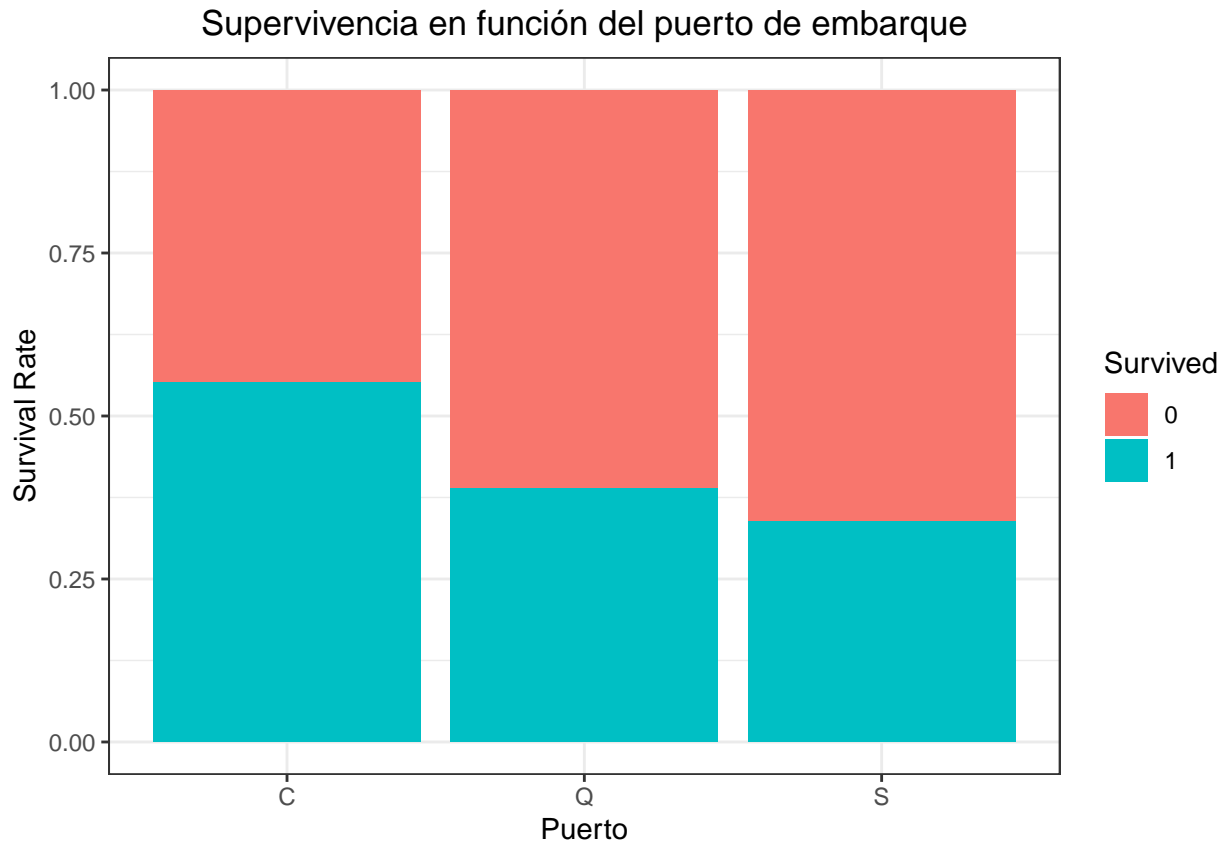
Creamos una nueva variable que nos indica si la persona es una madre o no. Esta variable se obtiene de las observaciones que cumplen la condición de tener el título 'Mrs' y que la variable Parch que indica el número de hijos o padres sea mayor que 0. Visualizamos la relación de nuestra nueva variable 'IsMother' con la supervivencia.



### Trabajo con Embarqued

En esta variable nos encontramos varios valores perdidos que sustituimos por el valor más frecuente. Visualizamos la relación de Embarqued con la supervivencia.





### Trabajo con Cabina

Existen demasiados valores perdidos y no tenemos una forma buena de predecirlos, sin que produzcan demasiado ruido, por lo que vamos a ignorar esta variable.

### Trabajo con Fare

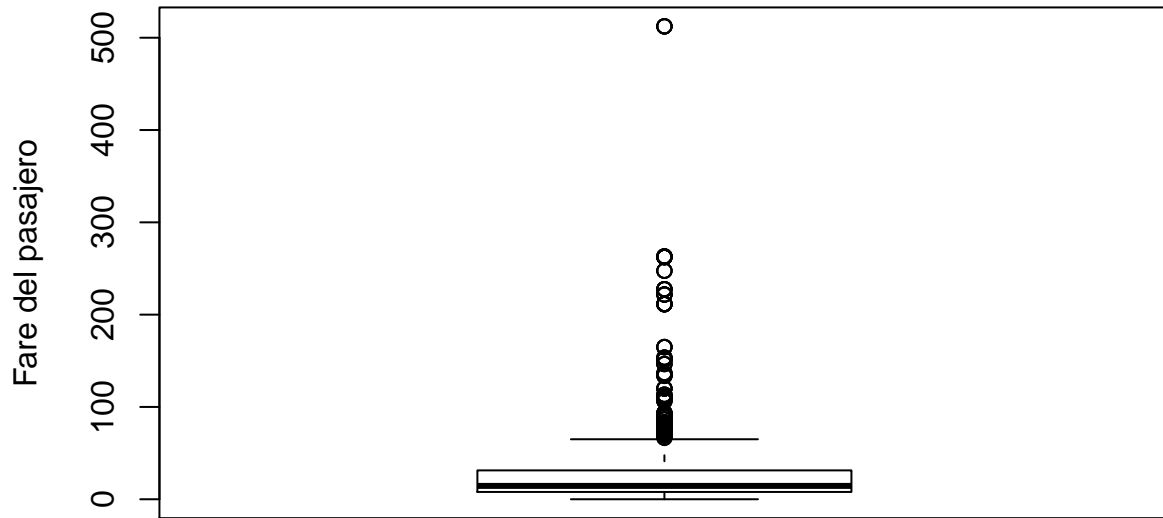
Existen varios valores perdidos que sustituimos por la mediana del 'Fare' en función de la clase, el sexo y el número de hijos. También existen ciertos pasajeros con Fare 0, pero los vamos a dejar como están.

Vemos el gráfico de caja de la variable y observamos que existen valores extremos que pueden afectar a la variable para solucionar esto vamos a agrupar Fare en categorías y añadimos los valores extremos a la última categoría. La agrupación se crea en la variable 'FareGroups' y visualizamos la relación de nuestras categorías con la supervivencia.

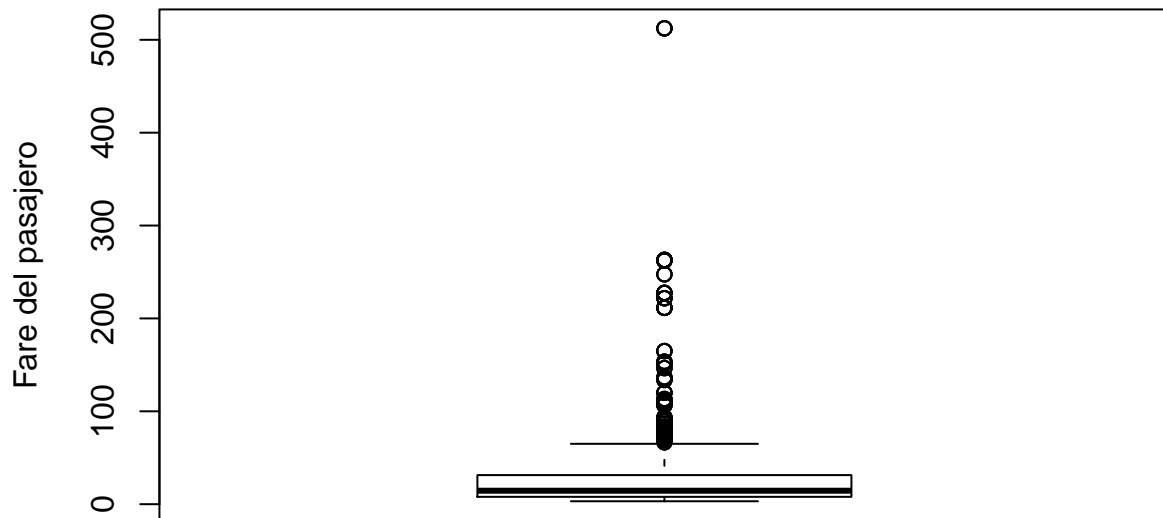
Otenemos una nueva variable con el Fare individual de cada persona, que obtenemos de dividir el Fare por el número de tickets iguales. Vemos que esta variable también está afectada por valores extremos, por lo que creamos una agrupación como en el caso de Fare(FareIndGroups) y visualizamos la relación de nuestros nuevos grupos con la supervivencia.

Como idea feliz vamos a probar a juntar las variables de Fare y sex, ya que hay diferencias significativas en la supervivencia de hombres y mujeres. Creamos una nueva variable 'FareIndividualBySex' que incrementa el Fare de las mujeres en 60, con esto conseguimos paliar un poco los efectos de los valores extremos, pero tendremos que analizar más adelante si es útil o no. También creamos una variable(FareIndividualBySexGroups) que agrupe estos nuevos Fares y visualizamos la relación de nuestros grupos con la supervivencia.

**grafico de caja de Fare**



**grafico de caja de Fare sin los ceros**



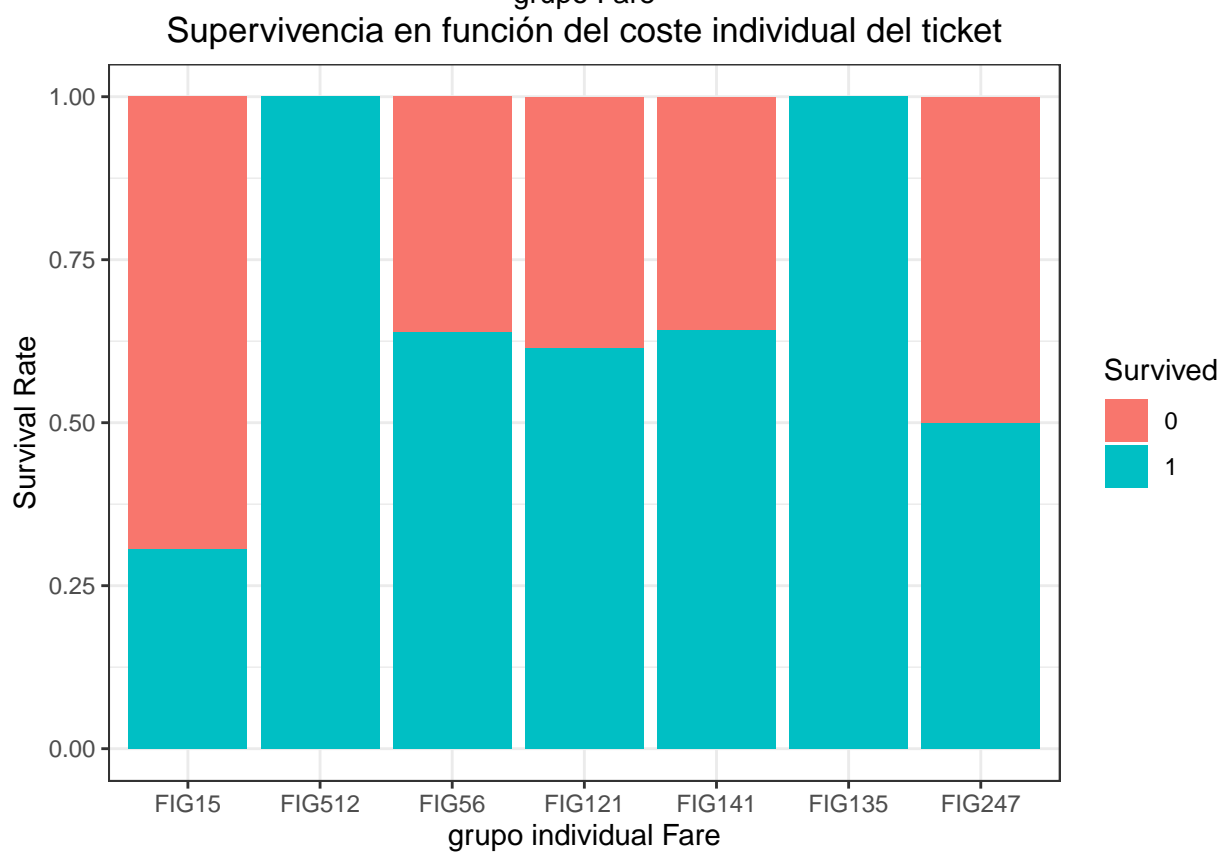
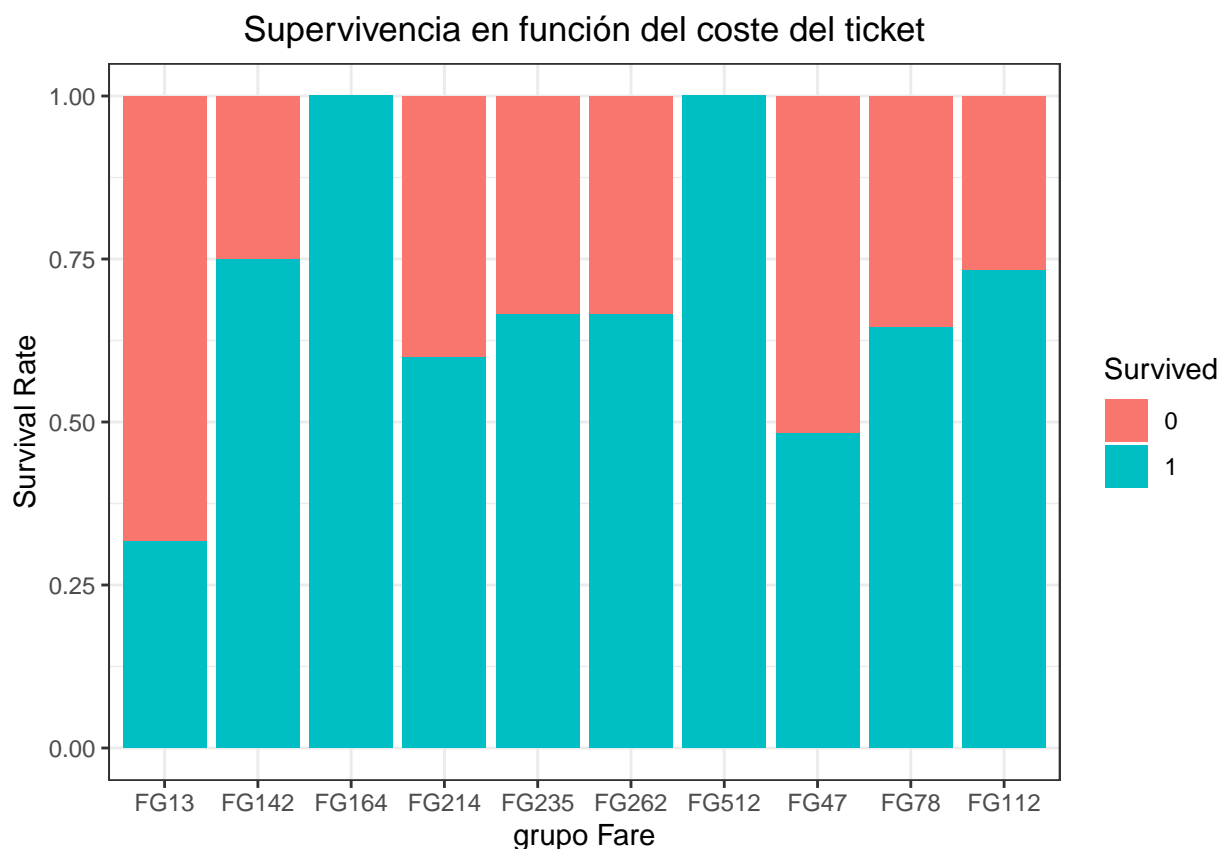


grafico de caja de FareIndividual

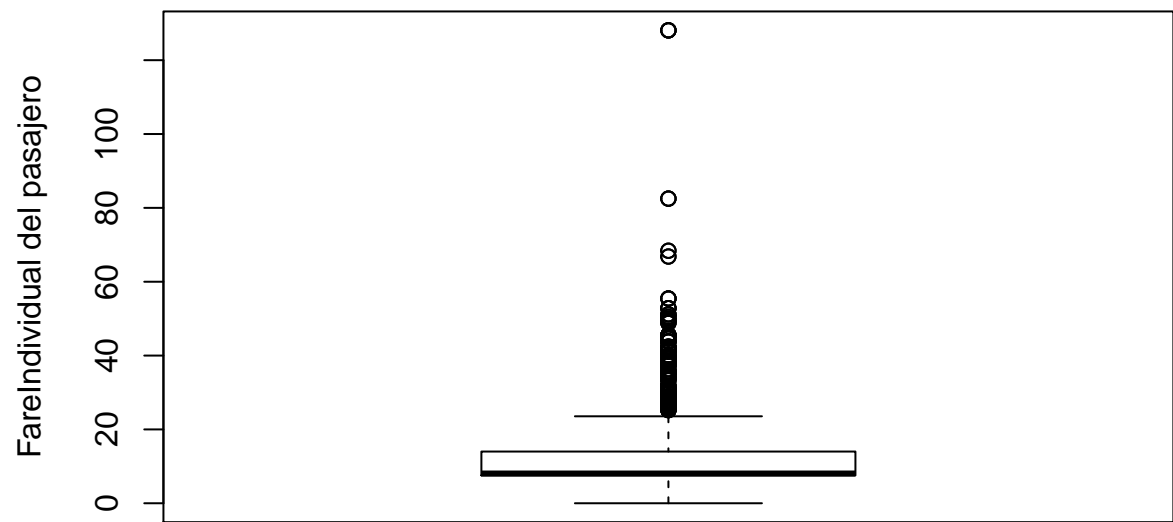
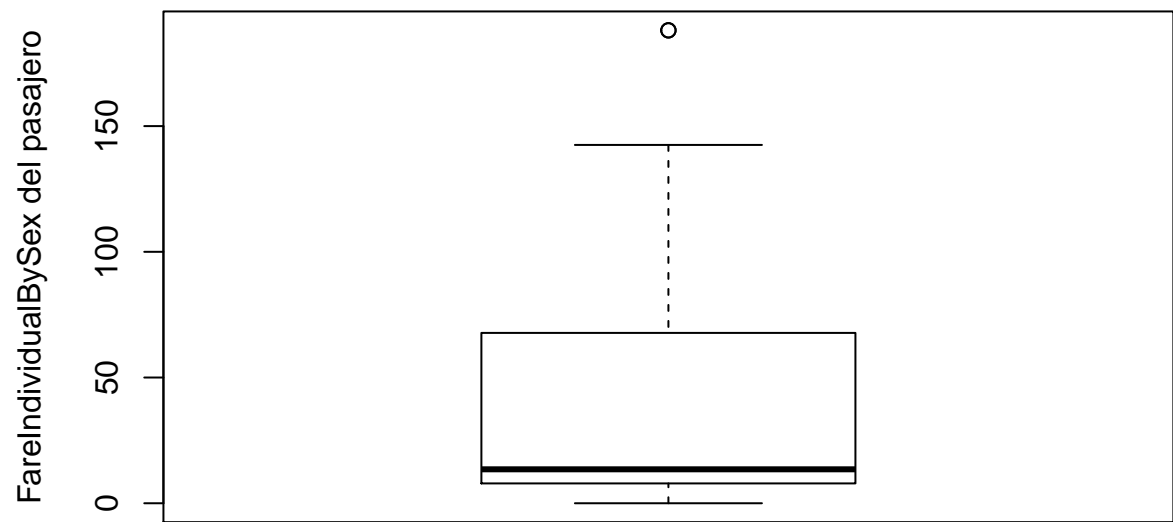
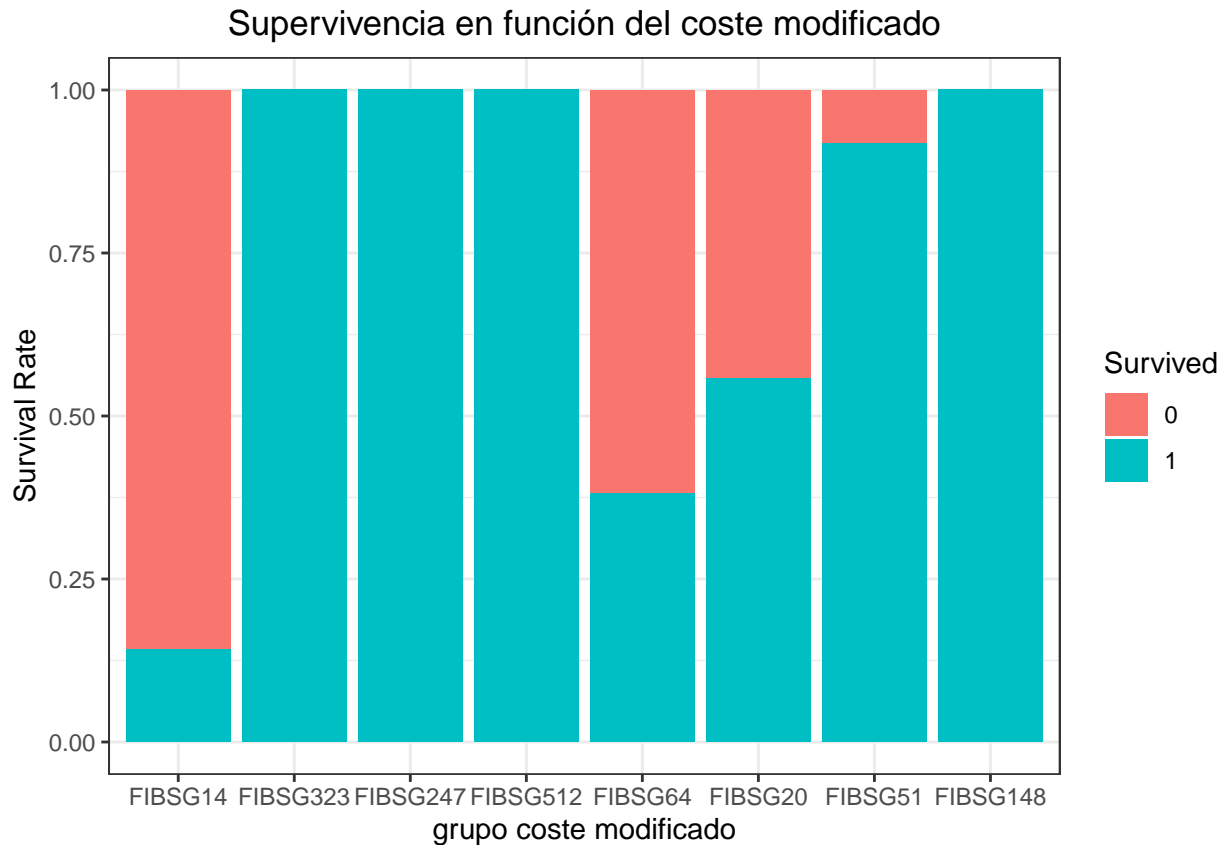


grafico de caja de FareIndividualBySexGroups





## Trabajo con Edad

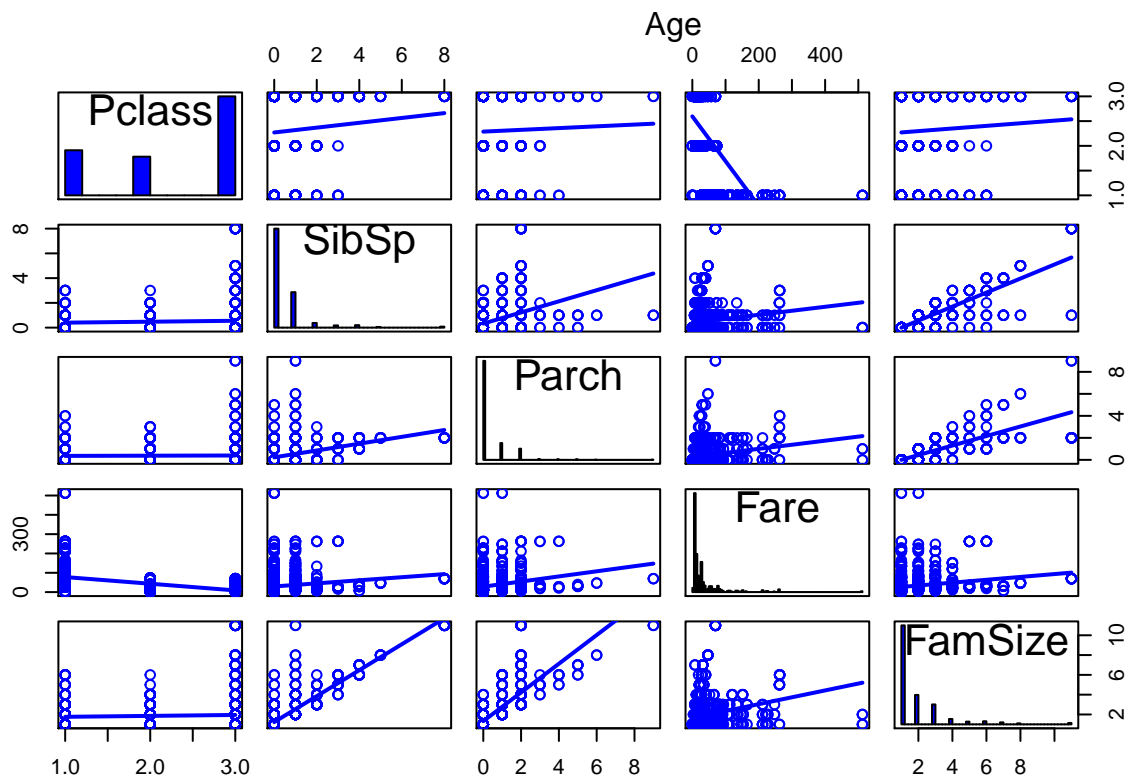
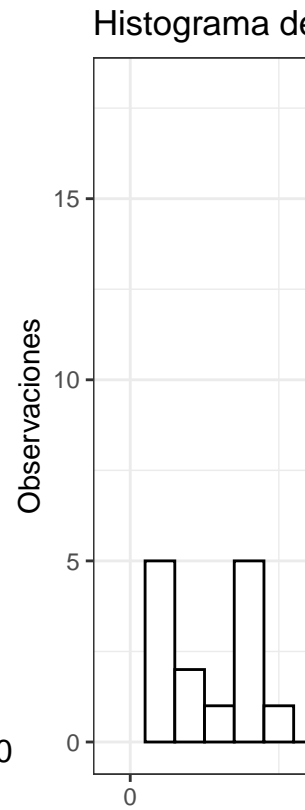
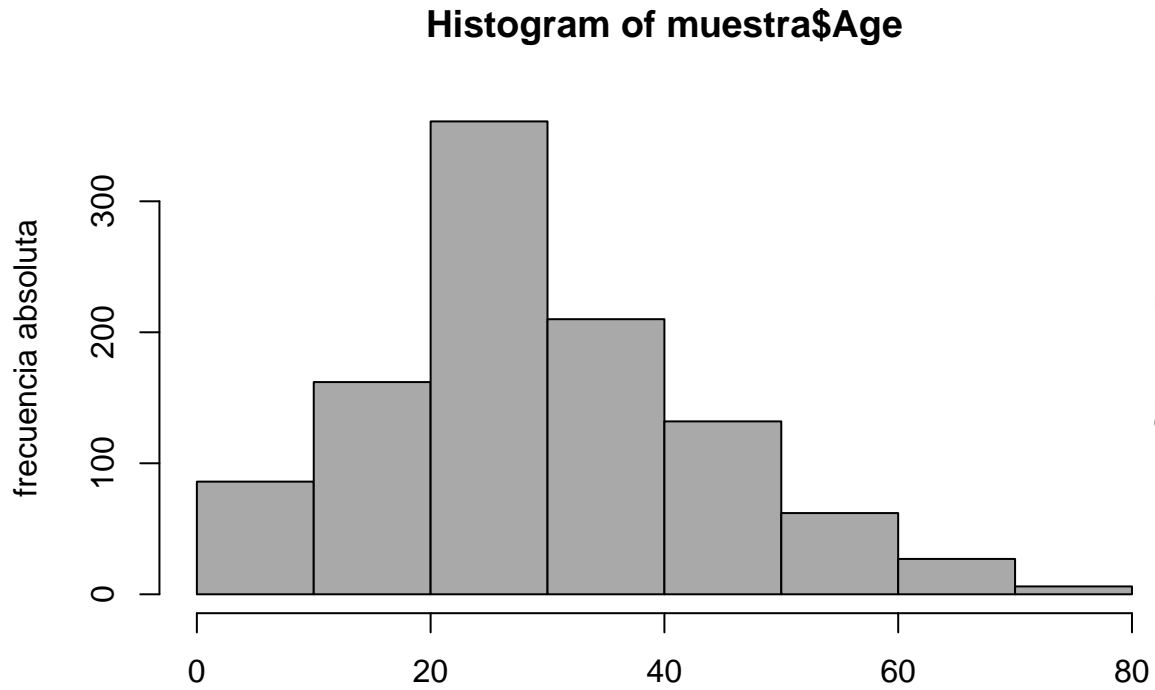
Mostramos el Histograma de la variable y vemos el número de agrupaciones óptimo para agrupar la variable. Creamos la variable 'ageGroups' con los grupos.

Primero vamos a limpiar los valores perdidos, esto lo vamos a hacer de varias maneras y luego estudiaremos cual es la mejor:

- El primer método de imputación será imputar por la edad media del título de la persona. Creamos la variable Ages.
- El segundo método es KNN con el que imputamos los valores perdidos de Age en función de las variables "Pclass", "Sex", "SibSp", "Parch", "Fare", "Embarked", "Titulo", 'FamSize'. Creamos la variable KNNAges
- El tercer método es por regresión. Primero vemos como se distribuye la edad visualizando su histograma con un mayor número de divisiones. Representamos la relación entre las variables cuantitativas y la edad. También mostramos la relación entre las variables cualitativas y la edad. Generamos un modelo para predecir el logaritmo de la edad (esto mejora la falta de normalidad de la variable Age) a partir de las variables Pclass, SibSp, Parch, Fare, Sex, Embarked, IsMother, LoneWolfs, FamSize, Titulo y usamos stepAIC para que seleccione el modelo que mejor funciona, partir de esas variables se va quedando con las variables más representativas. Visualizamos los gráficos de los residuos del modelo y creamos la variable RegresionAges para almacenar las edades imputadas.

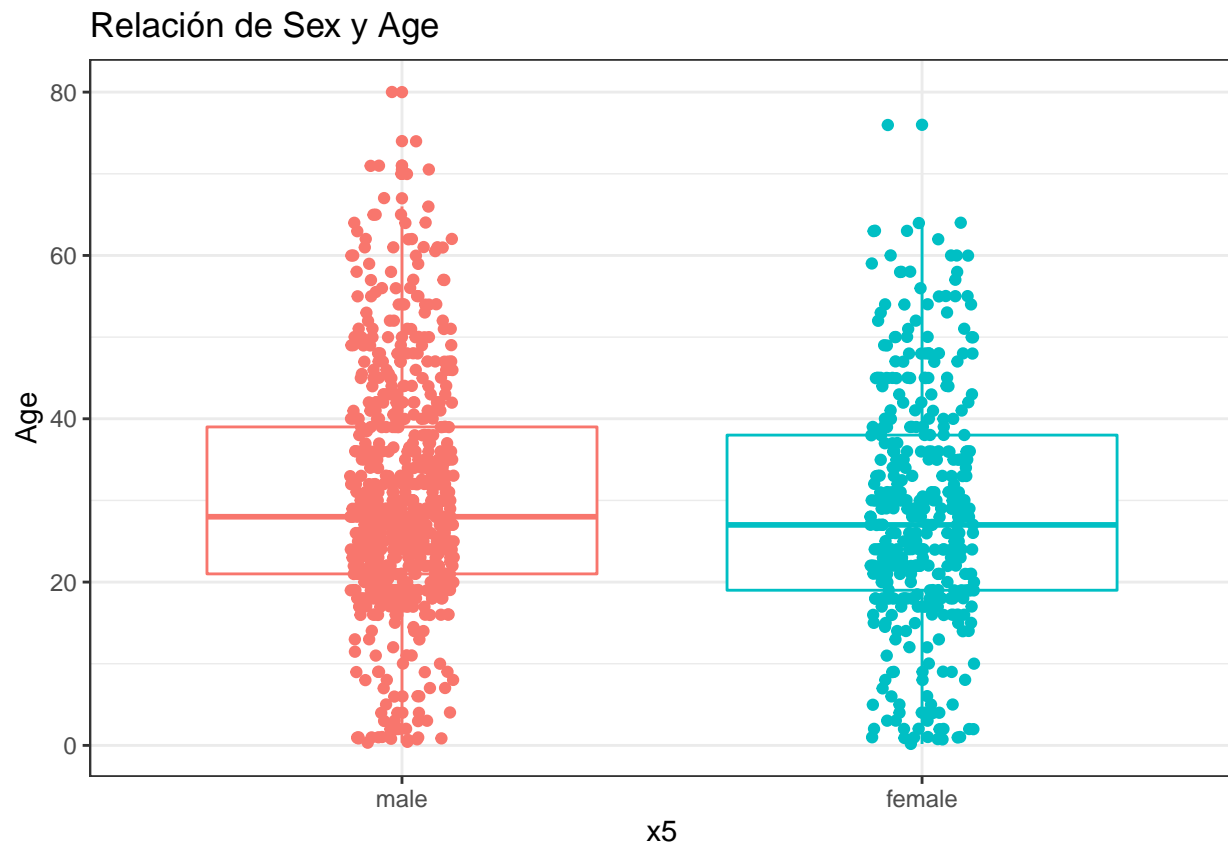
Visualizando el gráfico de caja de la variable observamos que hay algunos Outliers, para solucionarlo agrupamos las edades nuevamente tras la imputación de los valores perdidos y visualizamos la relación con supervivencia.

Creamos una nueva variable que nos indica si una persona menor de 16 años está viajando sola y visualizamos la relación con la supervivencia.



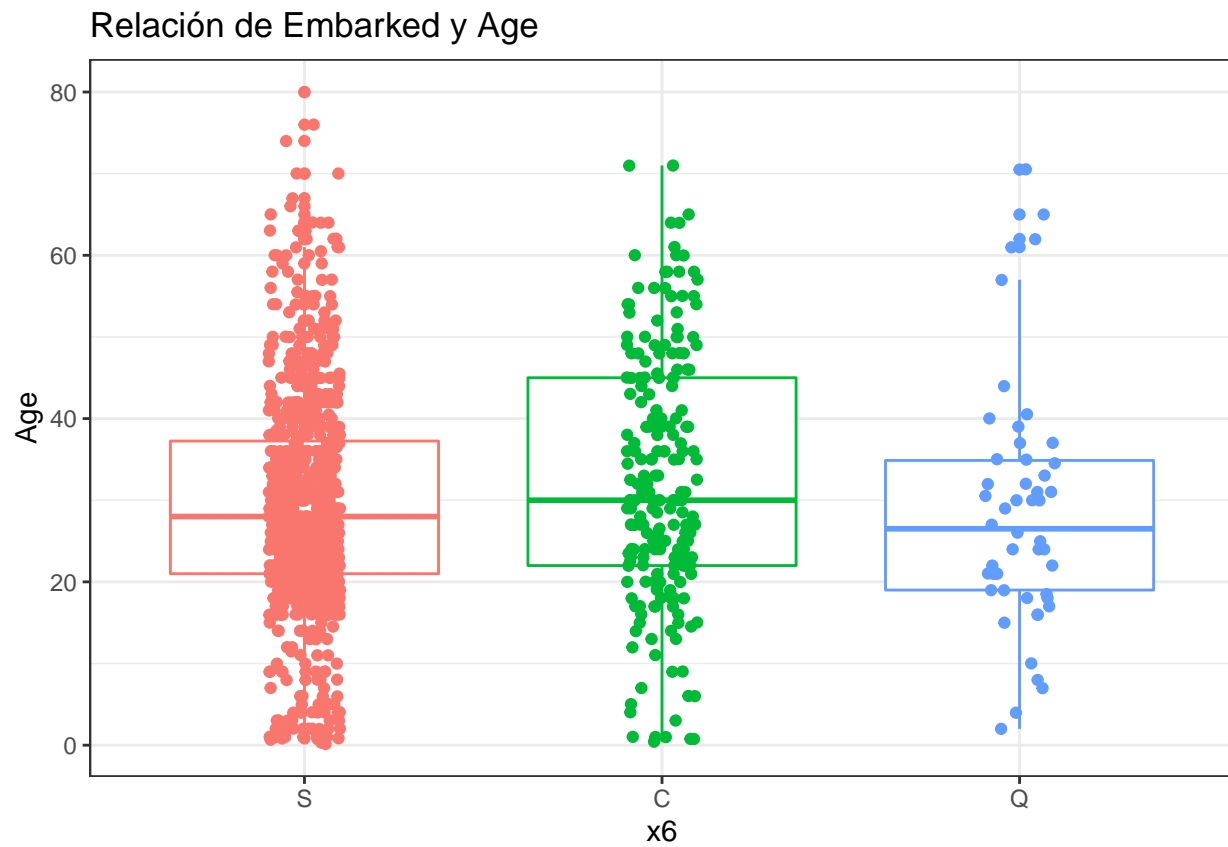
## Warning: Removed 263 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 263 rows containing missing values (geom\_point).



```
## Warning: Removed 263 rows containing non-finite values (stat_boxplot).
```

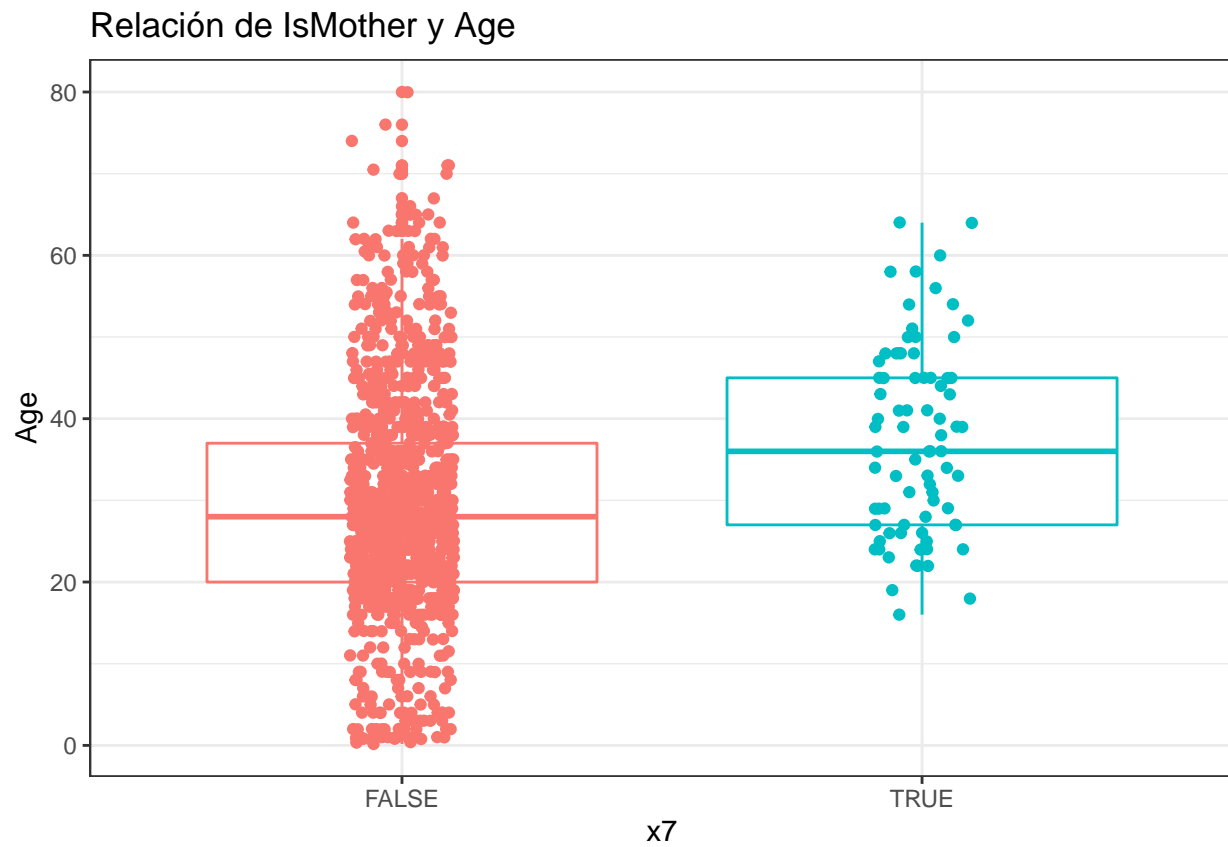
```
## Warning: Removed 263 rows containing missing values (geom_point).
```



```
## Warning: Removed 263 rows containing non-finite values (stat_boxplot).
```

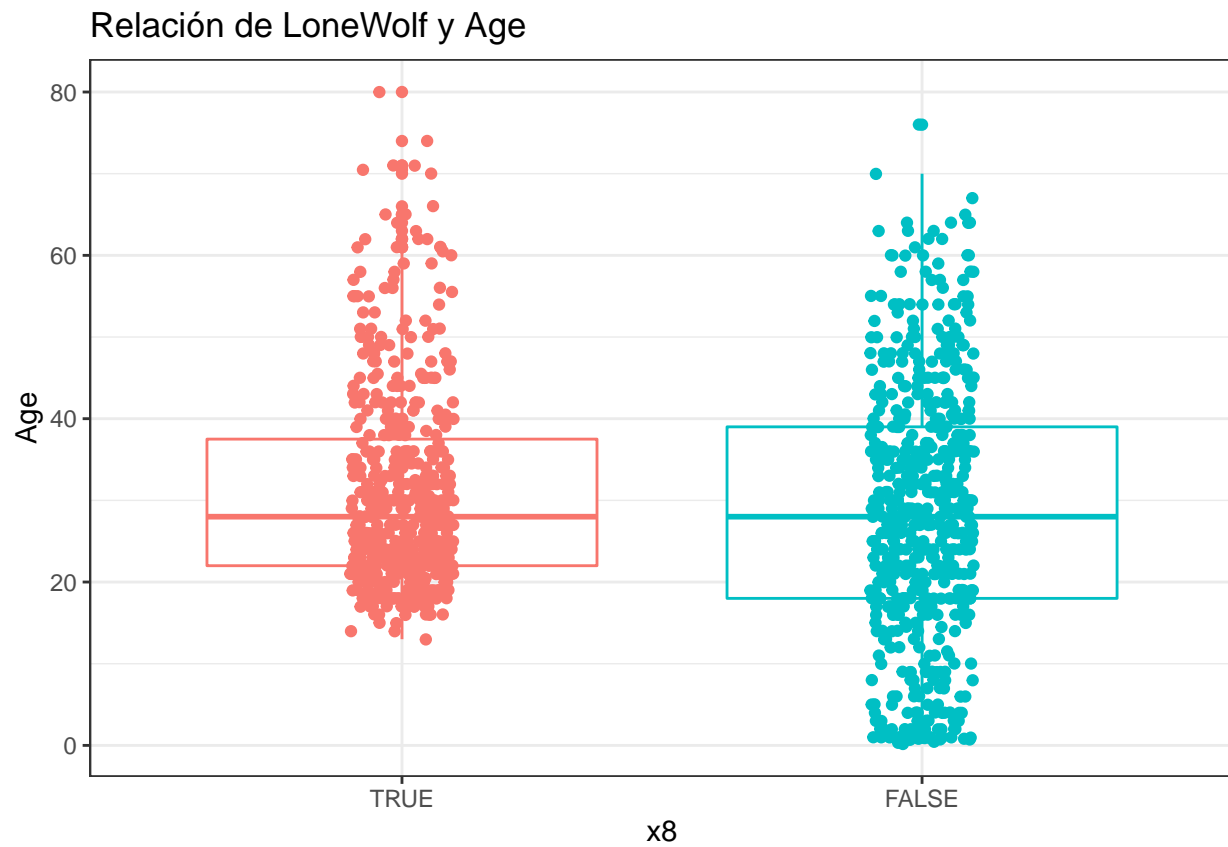
```
## Warning: Removed 263 rows containing missing values (geom_point).
```





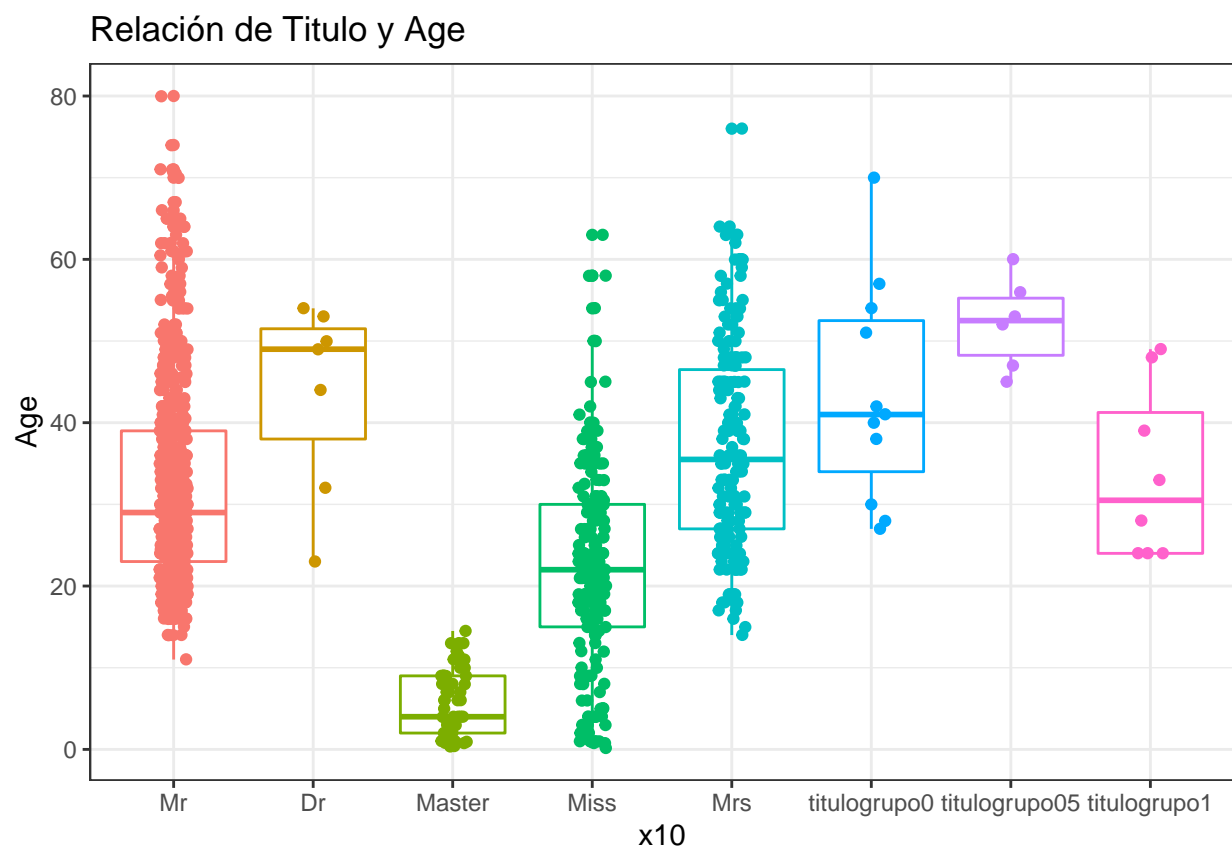
## Warning: Removed 263 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 263 rows containing missing values (geom\_point).



```
## Warning: Removed 263 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 263 rows containing missing values (geom_point).
```



```
## Warning: not plotting observations with leverage one:
## 101, 140, 150

## Warning: not plotting observations with leverage one:
## 101, 140, 150
```

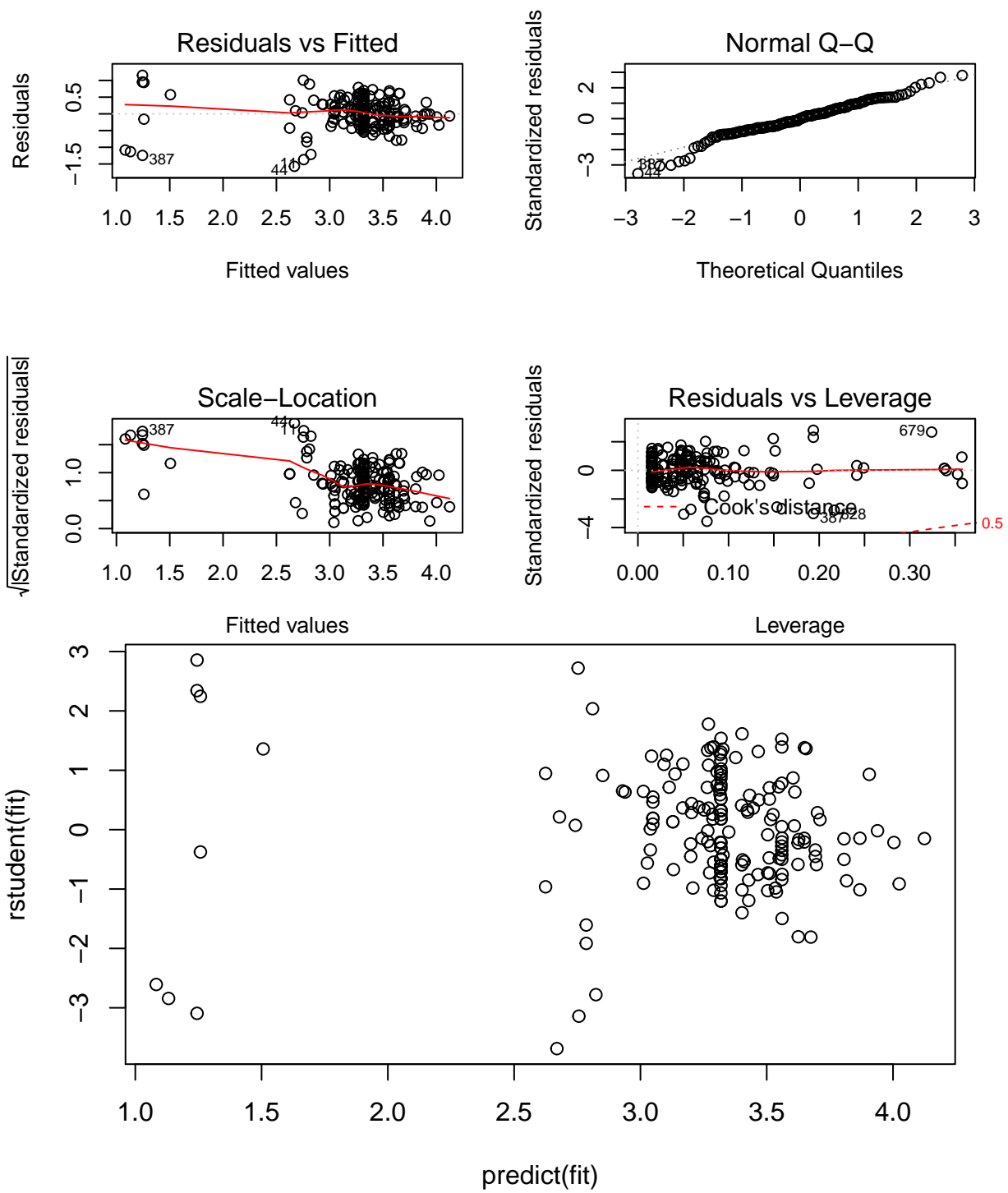
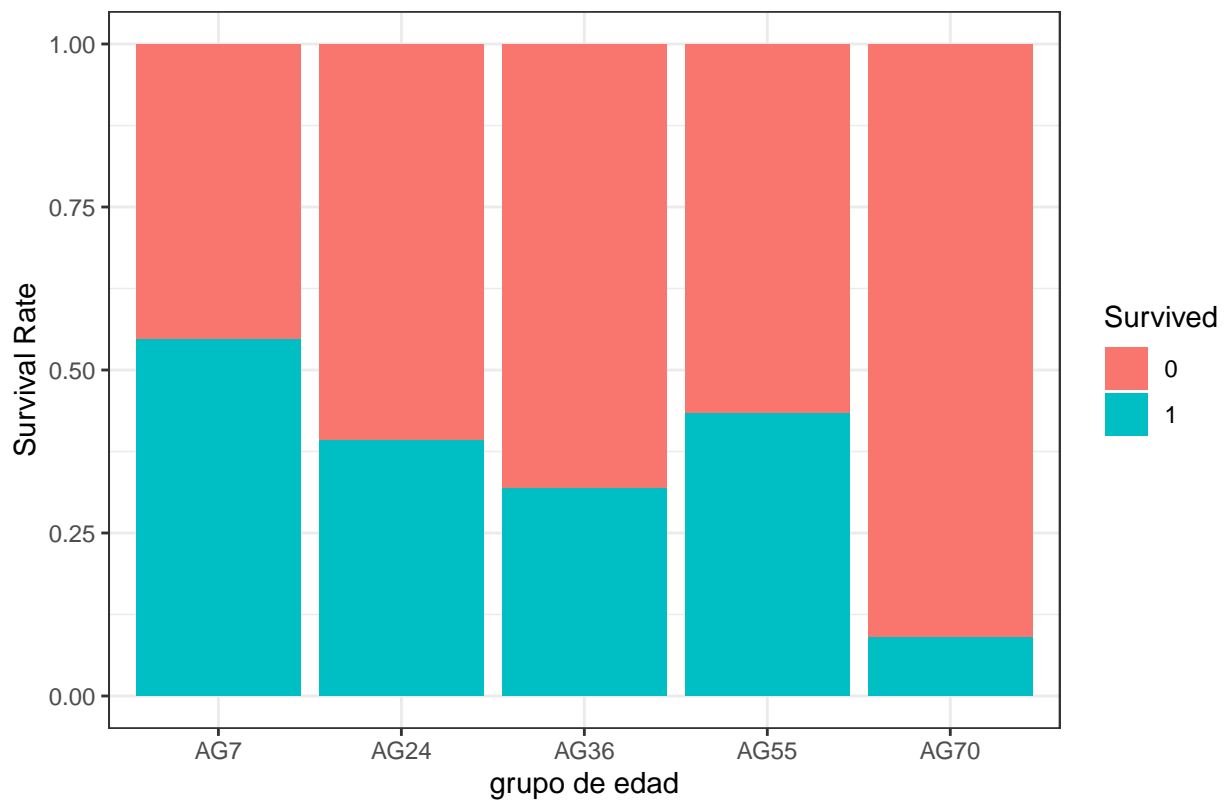
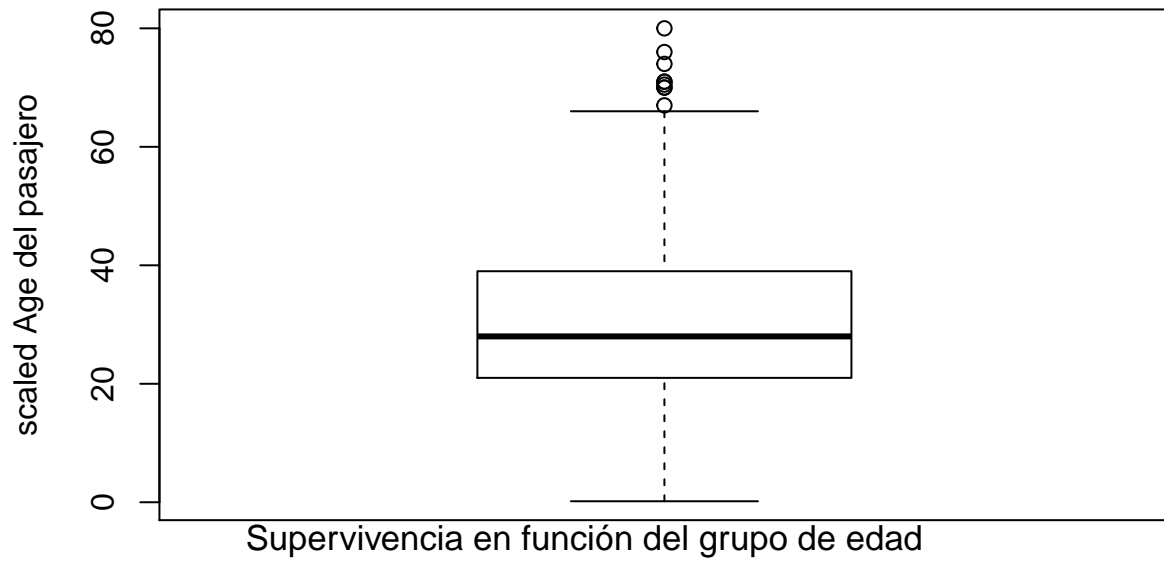
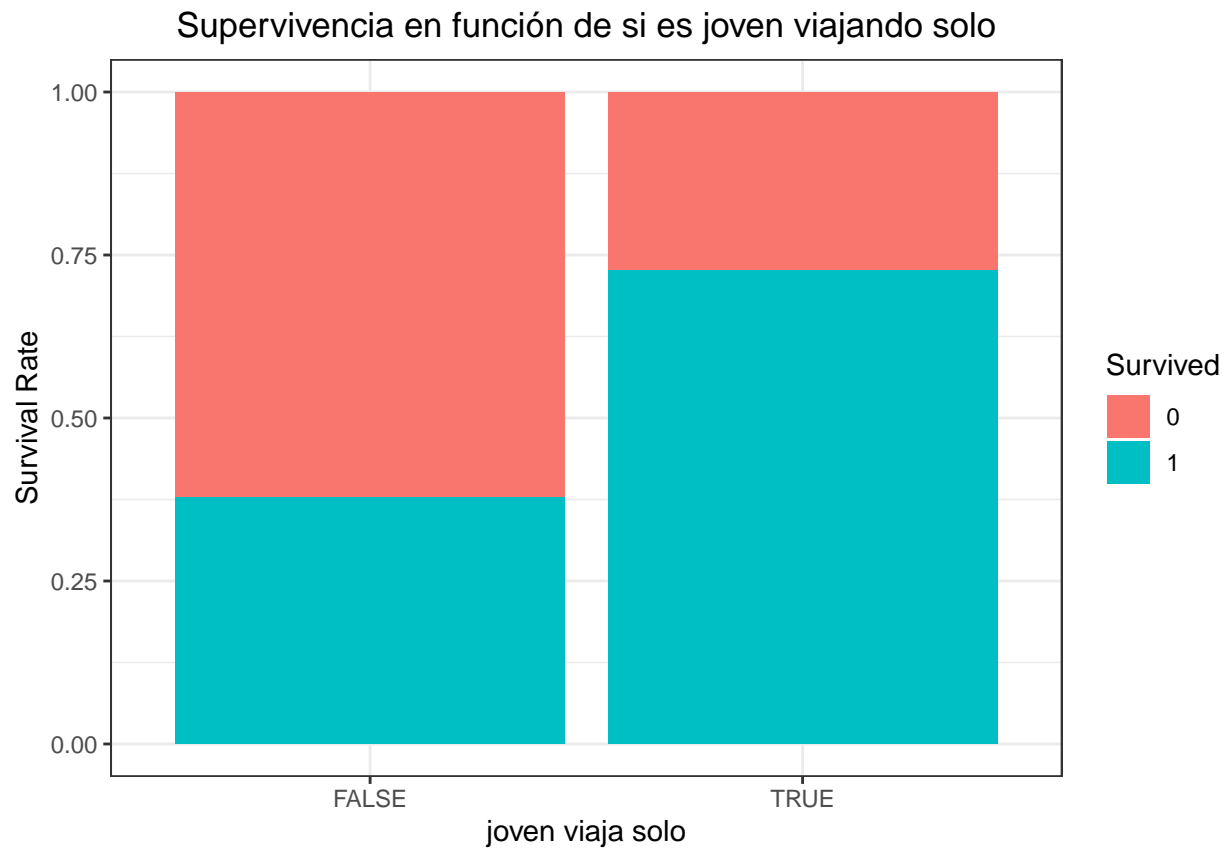


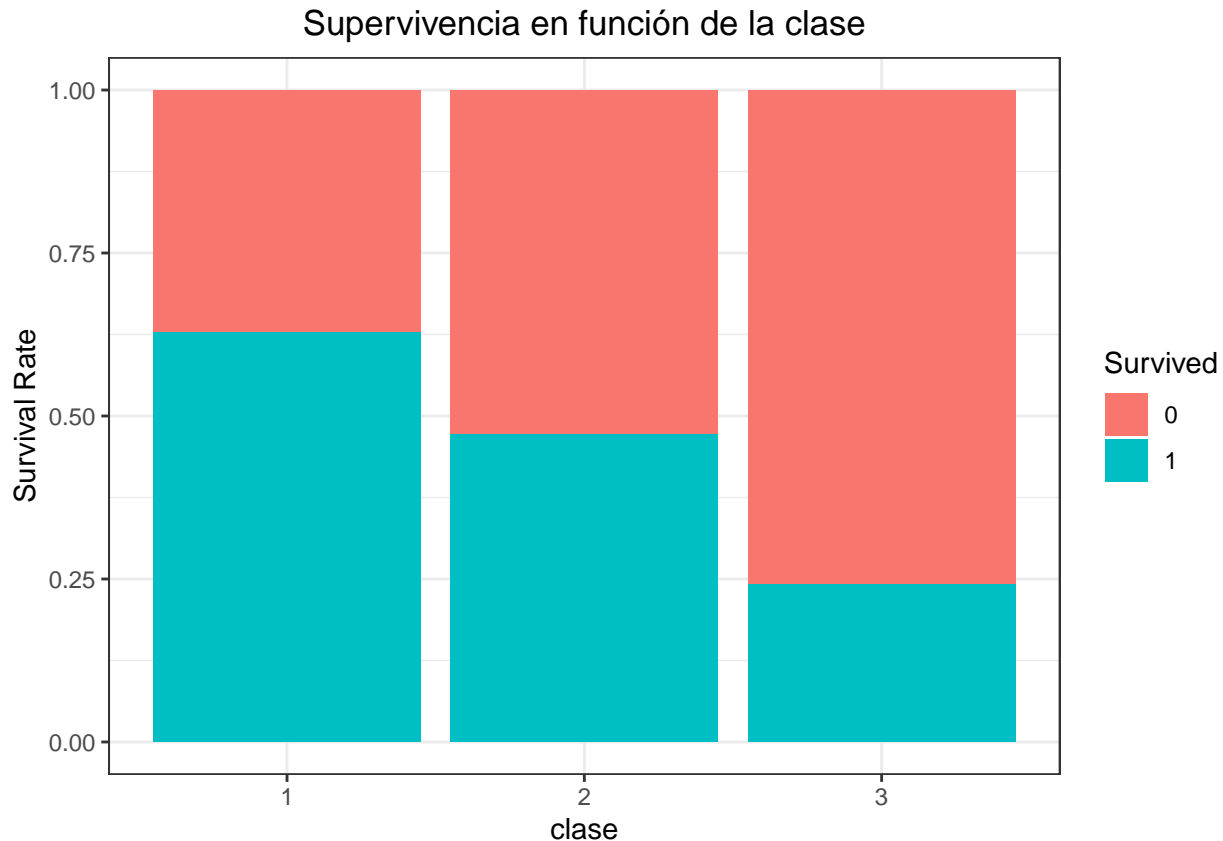
grafico de caja de Age





#### Trabajo con Pclass

Pclass está como integer pero realmente es una categórica, podemos crear una nueva variable tipo factor `PclassFactor`



### 3. Limpieza de los datos.

#### 3.1. ¿Los datos contienen ceros o elementos vacíos? ¿Cómo gestionarías cada uno de estos casos?

En la variable Ticket: Hemos tenido que corregir 4 casos de valores perdidos en nuestra nueva variable debido a que estos casos no tenían número de ticket y solo aparecía la palabra LINE. Como solo nos interesa para una variable que agrupa los tamaños de número de ticket le asignamos un 0 en dicha variable, en la variable original no hacemos nada porque no va a ser utilizada.

En la variable Embarked: Nos encontramos varios valores perdidos que sustituimos por el valor más frecuente.

En la variable Cabina: Existen demasiados valores perdidos y no tenemos una forma buena de predecirlos, sin que produzcan demasiado ruido, por lo que vamos a ignorar esta variable.

En la variable Fare: Existen ciertos pasajeros con Fare 0, pero los vamos a dejar como están.

En la variable Age: Limpiamos los valores perdidos, esto lo vamos a hacer de varias maneras y luego estudiaremos cual es la mejor: - El primer método de imputación será imputar por la edad media del título de la persona. Creamos la variable Ages. - El segundo método es KNN con el que imputamos los valores perdidos de Age en función de las variables "Pclass", "Sex", "SibSp", "Parch", "Fare", "Embarked", "Titulo", 'FamSize'. Creamos la variable KNNAges - El tercer método es por regresión. Primero vemos como se distribuye la edad visualizando su histograma con un mayor número de divisiones. Representamos la relación entre las variables cuantitativas y también mostramos la relación entre las variables cualitativas y la edad. Generamos un modelo para predecir el logaritmo de la edad (esto mejora la falta de normalidad de la variable Age) a partir de las variables Pclass, SibSp, Parch, Fare, Sex, Embarked, IsMother, LoneWolfs, FamSize, Titulo y

usamos `stepAIC` para que seleccione el modelo que mejor funciona, partir de esas variables se va quedando con las variables más representativas. Visualizamos los gráficos de los residuos del modelo y creamos la variable `RegresionAges` para almacenar las edades imputadas.

### 3.2. Identificación y tratamiento de valores extremos.

En la variable Fare: Para solucionar el problema de los valores extremos vamos agrupamos Fare en categorías y añadimos los valores extremos a la última categoría.

En la variable Age: Para solucionar el problema de los valores extremos agrupamos las edades e incluimos a los valores extremos en la última categoría.

#### 4. Análisis de los datos.

#### 4.1. Selección de los grupos de datos que se quieren analizar/comparar (planificación de los análisis a aplicar).

Inicialmente ponemos las variables en grupos en función de la relación que tienen por como se han creado:

"Fare", "FareGroups", "FareIndividual", "FareIndGroups", "FareIndividualBySex", "FareIndividualBySexGroups", "FarePw2", "Fare"

“Name”, “Titulo”, “FamilyIDTKGrouped”, “FamSize”, “TicketFreq”, “IsMother”

“SibSp”, “Parch”, “FamSize”, “IsMother”, “LoneWolfs”, “FamilyIDTKGrouped”, “youngTravelAl”

“Ticket”, “TicketFreq”, “TktNum”, “tamanoticket”, “FamilyIDTKGrouped”

“Age”, “ageGroupsByCut”, “Ages”, “KNNAges”, “RegresionAges”, “LoneWolfs”, “AgePw2”, “AgePw3”, “youngTravelAl”

“Sex”, “FareIndividualBySex”, “FareIndividualBySexGroups”

“Pclass”

“Embarked”

Hemos probado diferentes métodos de análisis para evaluar la eficacia de las variables:

## Análisis de las variables por métodos de filtrado

Este análisis se basa en contrastar la correlación de la variables, estudiar por contrastes de hipótesis si la variables independientes tienen alguna variable que es dependiente entre las variables explicativas o viendo la `information.gain` de las variables(paquete `FSelector`).

Vamos a analizar la asociacion entre variables de nuestros grupos y entre las variables y Survived:

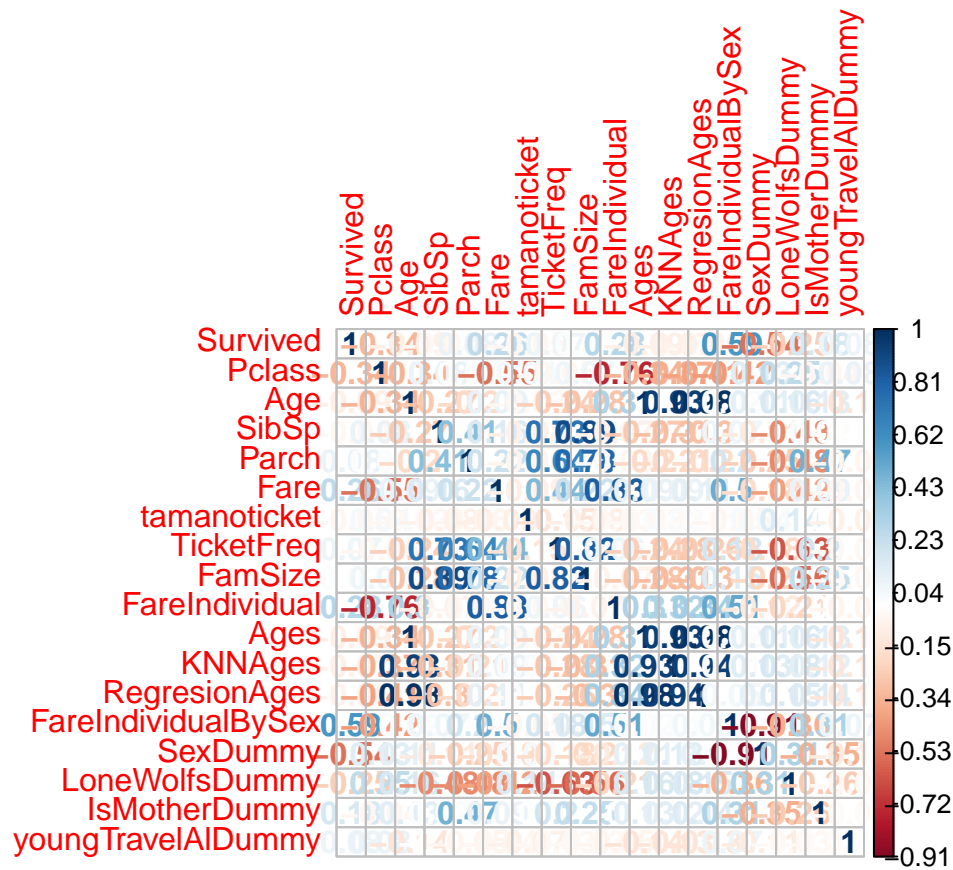
Mejor relacion con survived por grupos: "FareIndividualBySex","FareIndividualBySexGroups","FareIndividual"  
 "IsMother","TicketFreq" "LoneWolfs","IsMother" "TicketFreq" "LoneWolfs","ageGroupsByCut","Age"  
 "FareIndividualBySex","FareIndividualBySexGroups","Sex"

Grupos de variables interesantes tras este análisis: "FareIndividualBySex", "IsMother", "LoneWolfs", "TicketFreq", "ageGroupsBySex", "FareIndividualBySexGroups", "IsMother", "LoneWolfs", "TicketFreq", "ageGroupsByCut", "Embarked", "Pclass", "FarePw2", "FamilyIDTKGrouped", "Titulo", "FareIndividual", "IsMother", "LoneWolfs", "youngTravelAl", "Age", "Embarked", "Pclass"

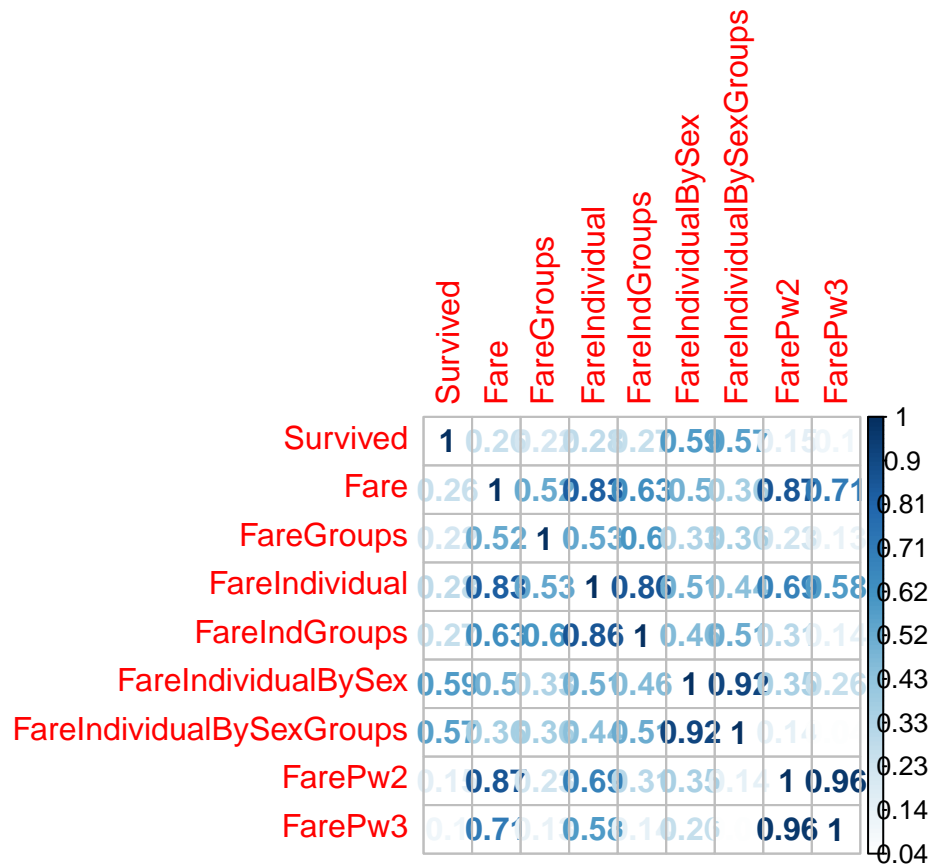
Mostrámos gráficamente las relaciones existentes

```
## [1] "Mostramos matriz de correlaciones entre las variables numéricas independientes"
```

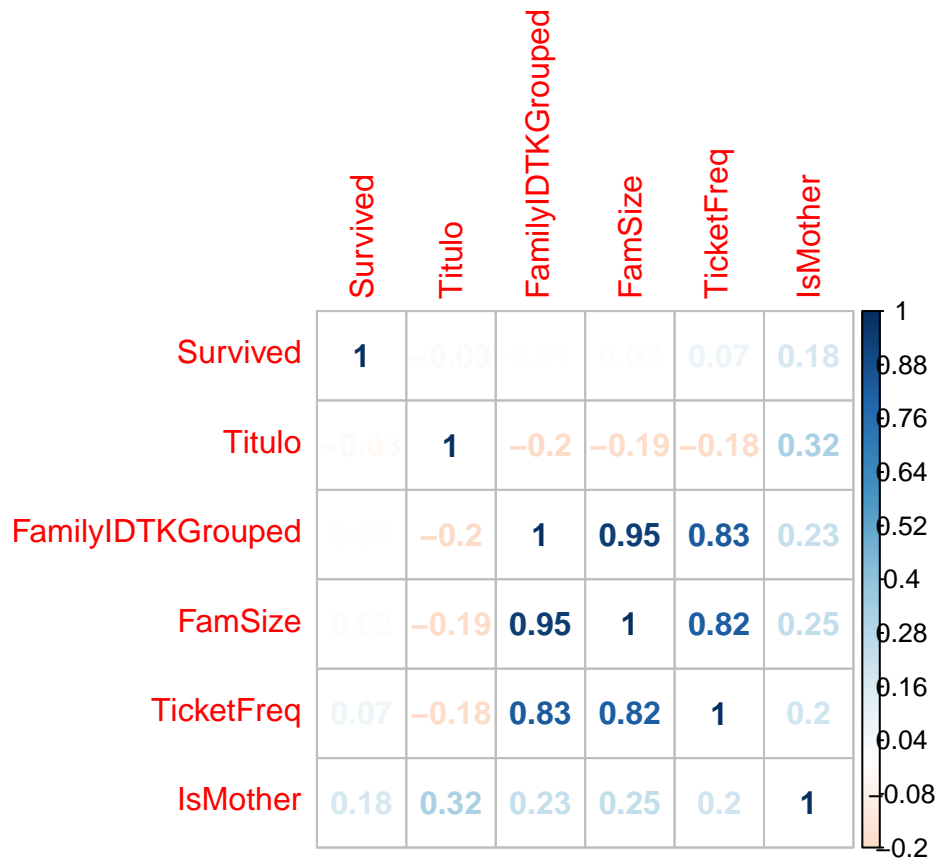




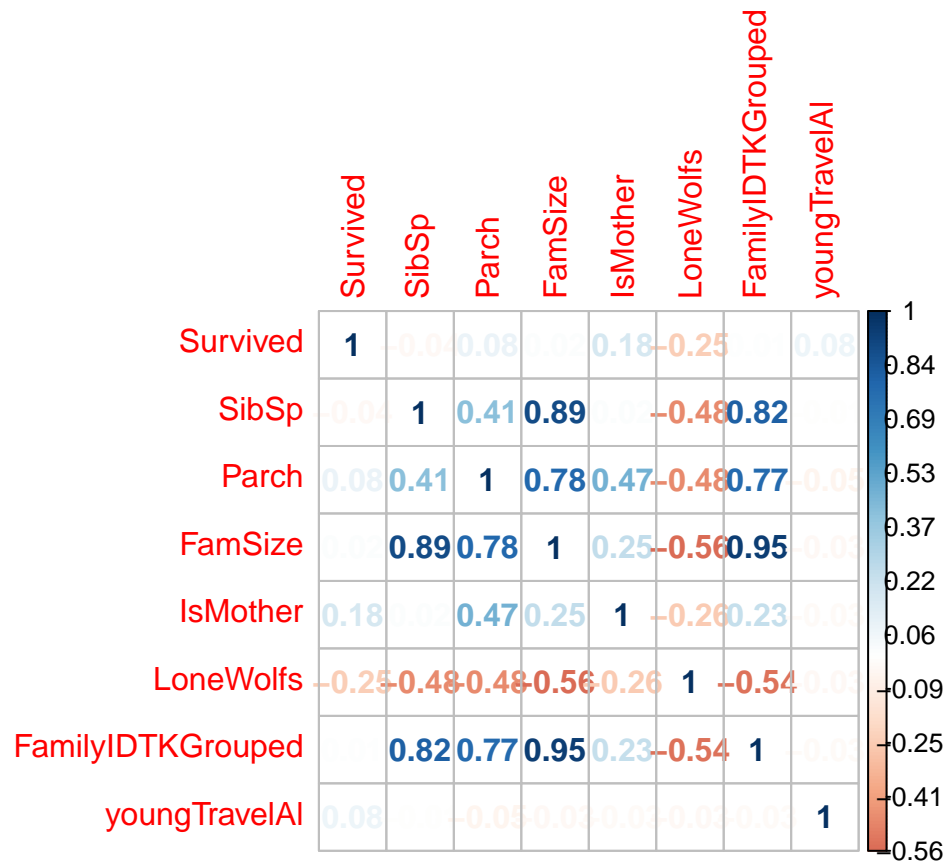
```
## [1] "Mostramos matriz de correlaciones entre las variables del grupo1"
```



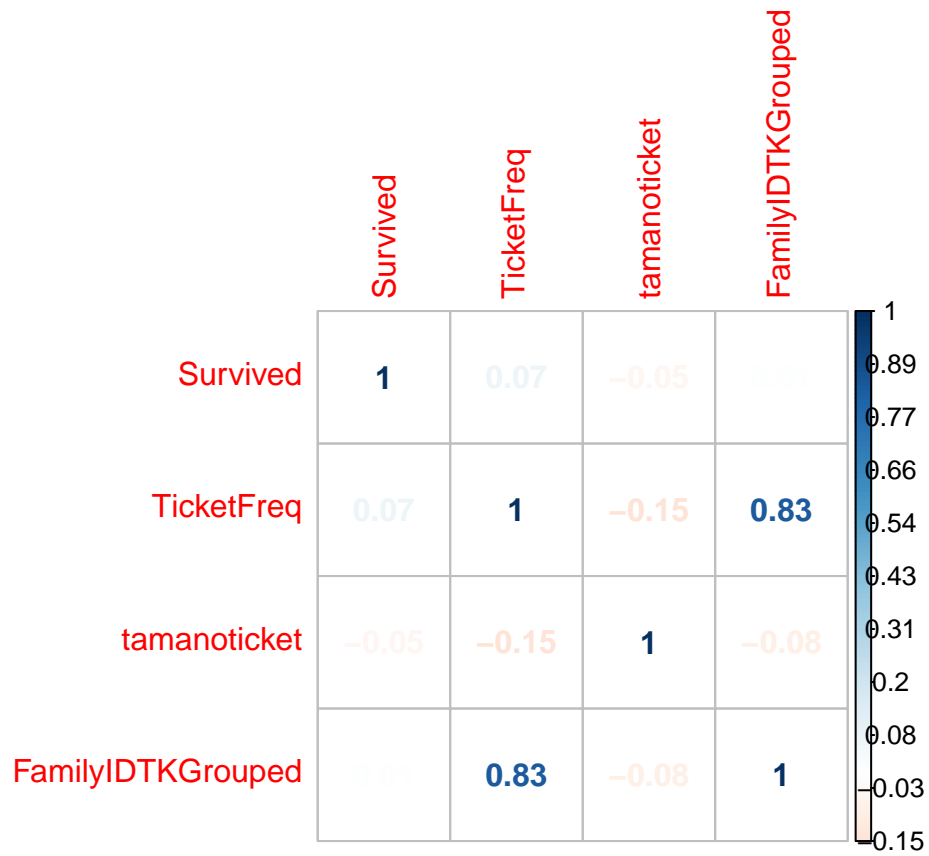
## [1] "Mostramos matriz de correlaciones entre las variables del grupo2"



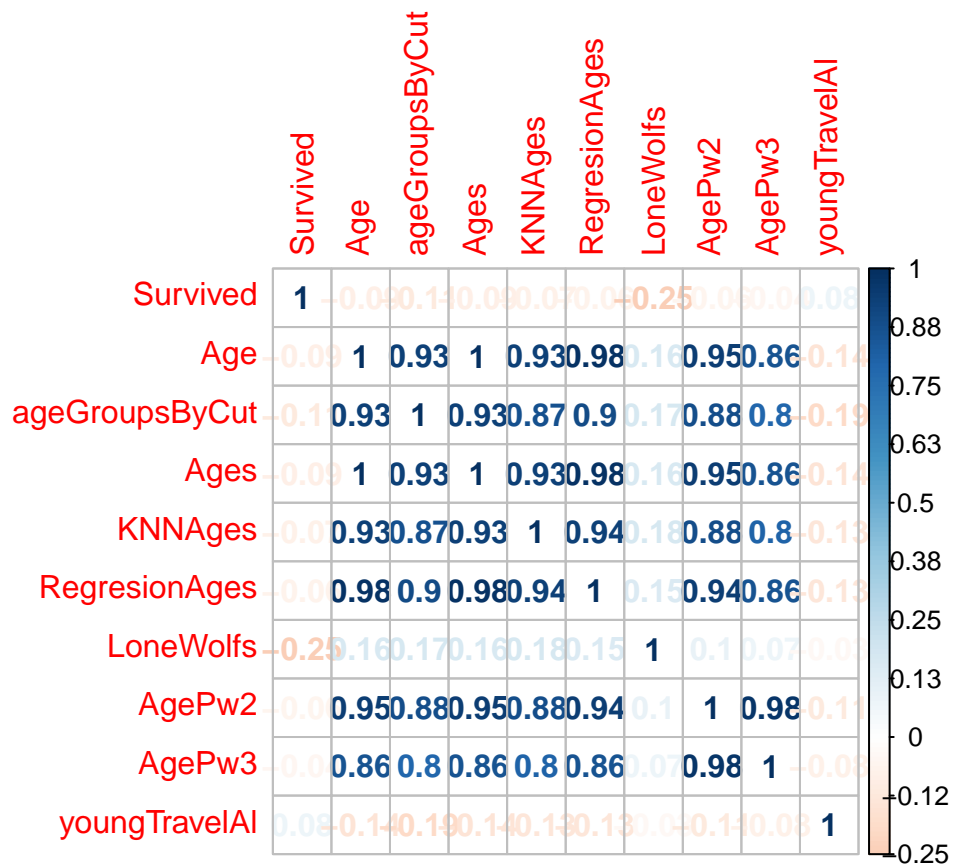
```
## [1] "Mostramos matriz de correlaciones entre las variables del grupo3"
```



```
## [1] "Mostramos matriz de correlaciones entre las variables del grupo4"
```



```
## [1] "Mostramos matriz de correlaciones entre las variables del grupo5"
```



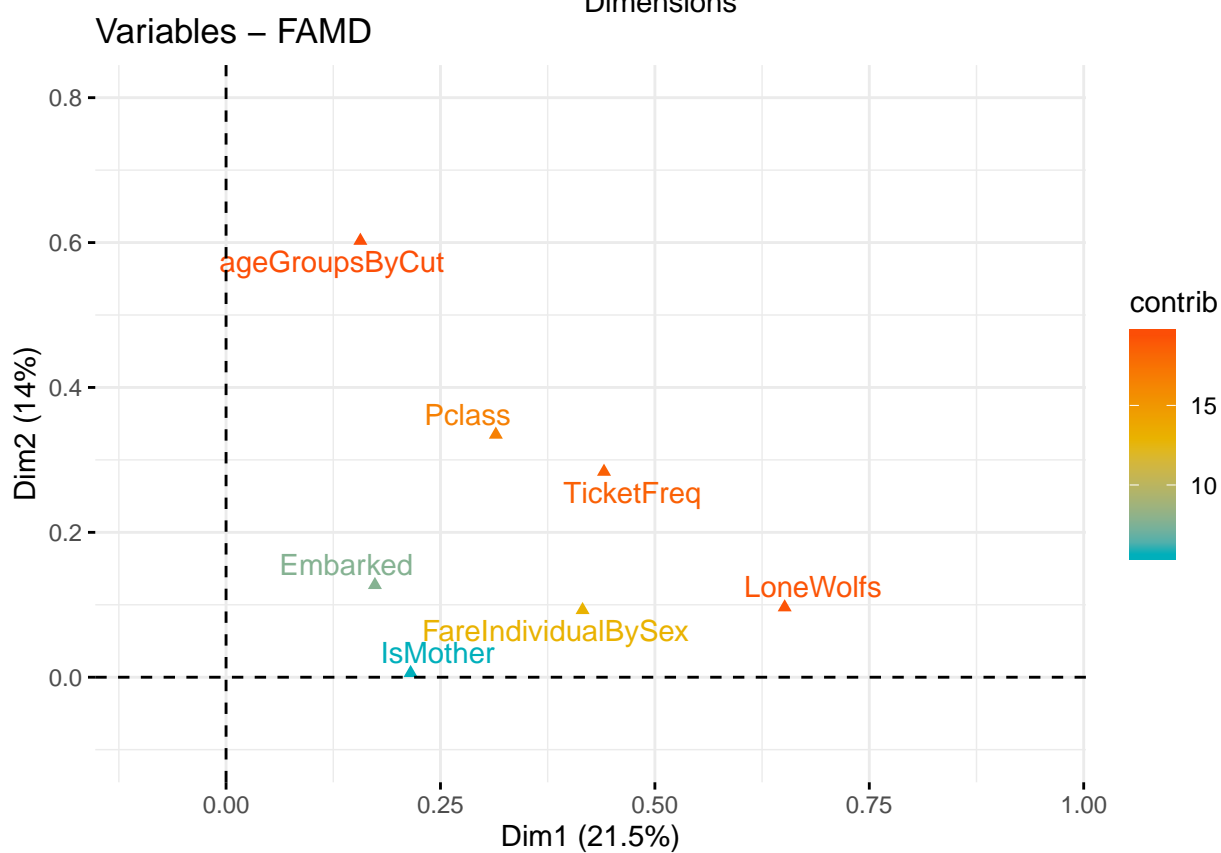
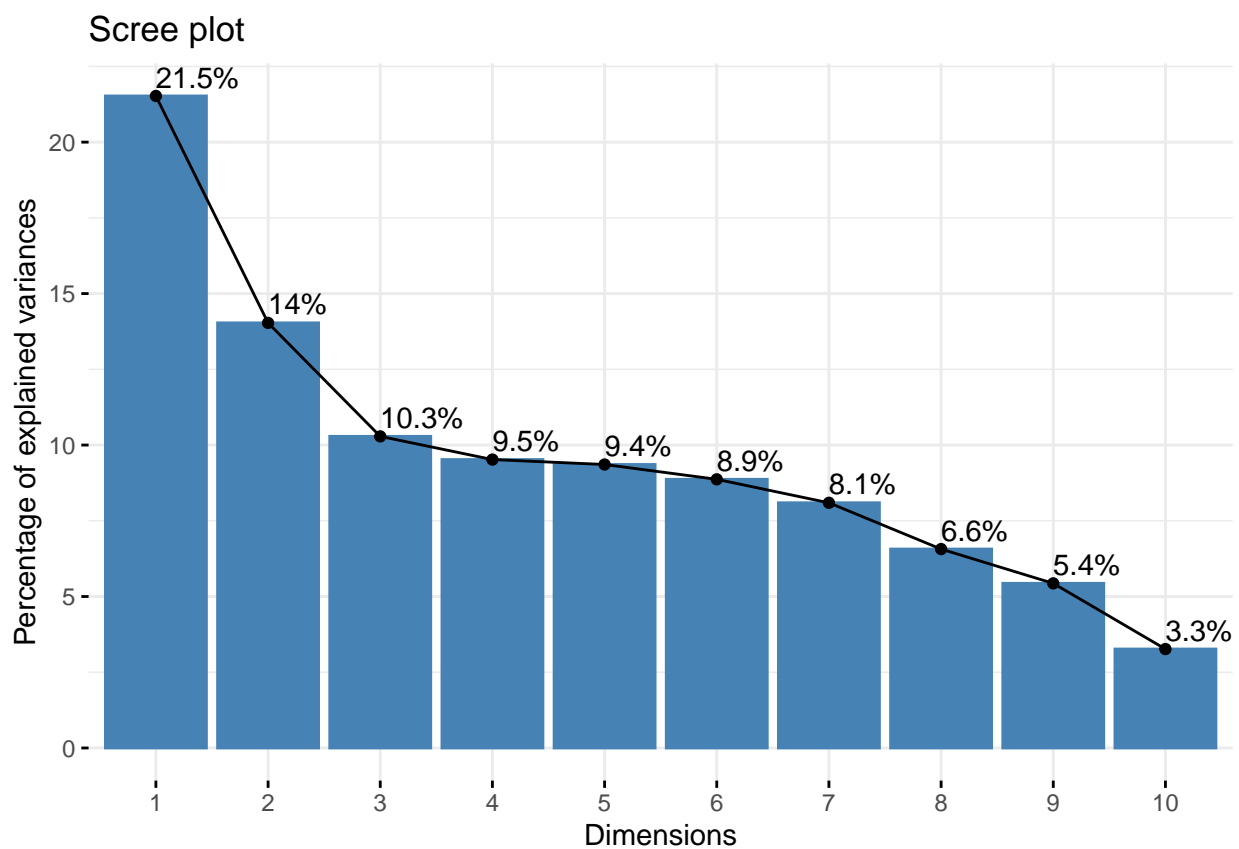
## [1] "Mostramos matriz de correlaciones entre las variables del grupo6"



## Análisis de las variables por análisis de factores

El análisis de factores es similar al análisis de componentes principales, le aportamos un conjunto de variables y construye un nuevo conjunto de datos con nuevas variables que representan a los datos originales. Los conjuntos de datos obtenidos con este sistema no nos han dado buenos resultados-

```
##      percentage of variance cumulative percentage of variance
## comp 1      21.522923      21.52292
## comp 2      14.032248      35.55517
## comp 3      10.287000      45.84217
## comp 4       9.520746      55.36292
## comp 5       9.358165      64.72108
## comp 6       8.868513      73.58959
## comp 7       8.093699      81.68329
## comp 8       6.566345      88.24964
## comp 9       5.434094      93.68373
## comp 10      3.265686      96.94942
## comp 11      3.050581     100.00000
```





### Análisis de las variables por Forward Selection, Backward Selection, Stepwise Selection

En este método generamos un modelo de regresión logística y mediante stepAIC va optimizando el modelo y eliminando las variables poco significativas.

## Análisis de las variables por Backward Selection con paquete mlr

Este método nos selecciona las variables más importantes del dataset, vamos a utilizar Backward selection que parte de todas las variables y va eliminando las menos significativas.

Variables recomendadas por este método: "Pclass", "FareIndividualBySex", "ageGroupsByCut" "Pclass", "Titulo", "FamSize", "FareIndGroups", "FareIndividualBySex", "AgePw2" "Pclass", "KNNAgesScaled", "FareIndividualBySexScaled" "SibSp", "FareIndividualBySex", "AgePw3", "PclassScaled"

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]























[illegible]















[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]







[illegible]





[illegible]



[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]





[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]





[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]





[illegible]

[illegible]

[illegible]



[illegible]

[illegible]

[illegible]



[illegible]

[illegible]







[illegible]







[illegible]

[illegible]



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



[illegible]

[illegible]



















```

## [FeatSel-y] 75: mmce.test.mean=0.1952862; time: 0.0 min
## [FeatSel] Result: Pclass,SibSp,Titulo,FareInd... (5 bits)
## Features          : 5
## Performance       : mmce.test.mean=0.1750842
## Pclass, SibSp, Titulo, FareIndividual, KNNAges
##
## Path to optimum:
## - Features: 78 Init : Perf = 0.22222 Diff: NA *
## - Features: 77 Remove : FareIndividualBySex Perf = 0.18855 Diff: 0.03367 *
## - Features: 76 Remove : PassengerId Perf = 0.18519 Diff: 0.003367 *
## - Features: 75 Remove : FareIndGroups Perf = 0.17508 Diff: 0.010101 *
## - Features: 74 Remove : Fare Perf = 0.17508 Diff: 0 *
## - Features: 73 Remove : FIBSG20 Perf = 0.17508 Diff: 0 *
## - Features: 72 Remove : PclassFactor Perf = 0.17508 Diff: 0 *
## - Features: 71 Remove : FIBSG323 Perf = 0.17508 Diff: 0 *
## - Features: 70 Remove : Size3 Perf = 0.17508 Diff: 0 *
## - Features: 69 Remove : Dr Perf = 0.17508 Diff: 0 *
## - Features: 68 Remove : FIG141 Perf = 0.17508 Diff: 0 *
## - Features: 67 Remove : FG78 Perf = 0.17508 Diff: 0 *
## - Features: 66 Remove : FG13 Perf = 0.17508 Diff: 0 *
## - Features: 65 Remove : Size2 Perf = 0.17508 Diff: 0 *
## - Features: 64 Remove : Age Perf = 0.17508 Diff: 0 *
## - Features: 63 Remove : youngTravelAl Perf = 0.17508 Diff: 0 *
## - Features: 62 Remove : titulogrupo0 Perf = 0.17508 Diff: 0 *
## - Features: 61 Remove : FamSize Perf = 0.17508 Diff: 0 *
## - Features: 60 Remove : FIG121 Perf = 0.17508 Diff: 0 *
## - Features: 59 Remove : AG7 Perf = 0.17508 Diff: 0 *
## - Features: 58 Remove : FIG512 Perf = 0.17508 Diff: 0 *
## - Features: 57 Remove : C Perf = 0.17508 Diff: 0 *
## - Features: 56 Remove : FG164 Perf = 0.17508 Diff: 0 *
## - Features: 55 Remove : titulogrupo1 Perf = 0.17508 Diff: 0 *
## - Features: 54 Remove : FarePw2 Perf = 0.17508 Diff: 0 *
## - Features: 53 Remove : FareGroups Perf = 0.17508 Diff: 0 *
## - Features: 52 Remove : FIBSG51 Perf = 0.17508 Diff: 0 *
## - Features: 51 Remove : Ages Perf = 0.17508 Diff: 0 *
## - Features: 50 Remove : FG142 Perf = 0.17508 Diff: 0 *
## - Features: 49 Remove : AG36 Perf = 0.17508 Diff: 0 *
## - Features: 48 Remove : FG214 Perf = 0.17508 Diff: 0 *
## - Features: 47 Remove : FG47 Perf = 0.17508 Diff: 0 *
## - Features: 46 Remove : FIBSG148 Perf = 0.17508 Diff: 0 *
## - Features: 45 Remove : Size7 Perf = 0.17508 Diff: 0 *
## - Features: 44 Remove : FIG135 Perf = 0.17508 Diff: 0 *
## - Features: 43 Remove : Size5 Perf = 0.17508 Diff: 0 *
## - Features: 42 Remove : Size6 Perf = 0.17508 Diff: 0 *
## - Features: 41 Remove : Master Perf = 0.17508 Diff: 0 *
## - Features: 40 Remove : LoneWolfs Perf = 0.17508 Diff: 0 *
## - Features: 39 Remove : titulogrupo05 Perf = 0.17508 Diff: 0 *
## - Features: 38 Remove : AG70 Perf = 0.17508 Diff: 0 *
## - Features: 37 Remove : FIG247 Perf = 0.17508 Diff: 0 *
## - Features: 36 Remove : AG55 Perf = 0.17508 Diff: 0 *
## - Features: 35 Remove : FIG15 Perf = 0.17508 Diff: 0 *
## - Features: 34 Remove : AgePw2 Perf = 0.17508 Diff: 0 *
## - Features: 33 Remove : Alone Perf = 0.17508 Diff: 0 *
## - Features: 32 Remove : FG512 Perf = 0.17508 Diff: 0 *
## - Features: 31 Remove : FIBSG64 Perf = 0.17508 Diff: 0 *
## - Features: 30 Remove : FarePw3 Perf = 0.17508 Diff: 0 *
## - Features: 29 Remove : SizeBig Perf = 0.17508 Diff: 0 *
## - Features: 28 Remove : Miss Perf = 0.17508 Diff: 0 *

```

```

## - Features: 27 Remove : FIBSG247 Perf = 0.17508 Diff: 0 *
## - Features: 26 Remove : Size4 Perf = 0.17508 Diff: 0 *
## - Features: 25 Remove : Mr Perf = 0.17508 Diff: 0 *
## - Features: 24 Remove : Sex Perf = 0.17508 Diff: 0 *
## - Features: 23 Remove : Q Perf = 0.17508 Diff: 0 *
## - Features: 22 Remove : FIBSG14 Perf = 0.17508 Diff: 0 *
## - Features: 21 Remove : FG262 Perf = 0.17508 Diff: 0 *
## - Features: 20 Remove : Mrs Perf = 0.17508 Diff: 0 *
## - Features: 19 Remove : FareIndividualByS... Perf = 0.17508 Diff: 0 *
## - Features: 18 Remove : FIBSG512 Perf = 0.17508 Diff: 0 *
## - Features: 17 Remove : IsMother Perf = 0.17508 Diff: 0 *
## - Features: 16 Remove : S Perf = 0.17508 Diff: 0 *
## - Features: 15 Remove : Parch Perf = 0.17508 Diff: 0 *
## - Features: 14 Remove : AG24 Perf = 0.17508 Diff: 0 *
## - Features: 13 Remove : Embarked Perf = 0.17508 Diff: 0 *
## - Features: 12 Remove : FG112 Perf = 0.17508 Diff: 0 *
## - Features: 11 Remove : FIG56 Perf = 0.17508 Diff: 0 *
## - Features: 10 Remove : AgePw3 Perf = 0.17508 Diff: 0 *
## - Features: 9 Remove : TicketFreq Perf = 0.17508 Diff: 0 *
## - Features: 8 Remove : RegresionAges Perf = 0.17508 Diff: 0 *
## - Features: 7 Remove : tamanoticket Perf = 0.17508 Diff: 0 *
## - Features: 6 Remove : ageGroupsByCut Perf = 0.17508 Diff: 0 *
## - Features: 5 Remove : FG235 Perf = 0.17508 Diff: 0 *
##
## Stopped, because no improving feature was found.

```

## Análisis de las variables por Recursive Feature Elimination Method (RFE)

Este método es similar al anterior pero también obtenemos un listado con las variables ordenadas por importancia

Variables recomendadas por este método: “FareIndividualBySex”, “Titulo”, “FareIndividual”, “KNNAges”, “Sex” FareIndividualBySex, Titulo, KNNAgesScaled

Sin correlacion entre ellas: “FareIndividualBySex”, “Titulo”, “KNNAges”, ‘PclassScaled’, ‘FamilyIDTKGrouped’, ‘Embarked’, ‘Lone

```

##
## Recursive feature selection
##
## Outer resampling method: Cross-Validated (10 fold, repeated 5 times)
##
## Resampling performance over subset size:
##
## Variables Accuracy Kappa AccuracySD KappaSD Selected
##      1  0.7970 0.5645  0.04437 0.09760
##      2  0.7979 0.5661  0.04293 0.09539
##      3  0.7892 0.5416  0.04129 0.09212
##      4  0.7910 0.5503  0.04079 0.08814
##      5  0.7928 0.5521  0.04267 0.09162
##      6  0.7936 0.5517  0.04553 0.09866
##      7  0.7970 0.5606  0.04626 0.09899
##      8  0.8015 0.5690  0.04759 0.10356
##      9  0.7919 0.5508  0.04546 0.09728
##     10  0.7922 0.5528  0.04937 0.10420
##     47  0.8045 0.5805  0.04016 0.08628      *
##
## The top 5 variables (out of 47):
##      FareIndividualBySex, FareIndividualBySexScaled, Titulo, KNNAgesScaled, KNNAges

```

```

## Overall
## FareIndividualBySex 13.600949
## FareIndividualBySexScaled 13.582484
## Titulo 11.363564
## SexDummy 8.709066
## KNNAgesScaled 8.703271
## KNNAges 8.636648
## FareIndividualScaled 8.454495
## FareIndividual 8.335605
## Sex 8.206359
## AgePw3 8.007264
## Ages 7.957815
## PclassScaled 7.956500
## AgePw2 7.909015
## FarePw2 7.904552
## FareScaled 7.886852
## Age 7.867687
## Pclass 7.855544
## Fare 7.839743
## AgesScaled 7.827489
## AgeScaled 7.809420
## FarePw3 7.754604
## PclassFactor 7.513338
## FareIndividualBySexGroups 7.468243
## RegresionAgesScaled 7.398841
## RegresionAges 7.226660
## FamilyIDTKGrouped 6.101545
## TicketFreq 5.329217
## TicketFreqScaled 5.297692
## Embarked 4.907116
## FamSize 4.225986
## FamSizeScaled 4.124099
## ageGroupsByCut 3.835092
## FareIndGroups 3.317044
## SibSp 3.287760
## SibSpScaled 3.198566
## FareGroups 3.127860
## LoneWolfsDummy 2.815112
## LoneWolfs 2.677530
## PassengerId 1.720031
## youngTravelAl 1.499671
## IsMother 1.492384
## IsMotherDummy 1.479338
## tamanoticket 1.476044
## youngTravelAlDummy 1.449961
## tamanoticketScaled 1.343119
## ParchScaled 1.131183
## Parch 1.066165

```

## Análisis de las variables RandomForest

En este método utilizamos random forest, que genera árboles de decisión aleatorios y va comparando los resultados obtenidos para filtrar las variables más importantes.

Variables recomendadas por este método: 'FareIndividualBySex', 'Titulo', 'Mr', 'KNNAges', 'FIBSG14', 'FamilyIDTKGrouped', 'Pclass', 'FareIndividualBySex', 'Titulo', 'KNNAges', 'FamilyIDTKGrouped', 'PclassFactor', 'tamanoticket', 'Embarked', 'LoneWolfsDummy', 'PassengerId', 'youngTravelAl', 'IsMother', 'IsMotherDummy', 'tamanoticketScaled', 'ParchScaled', 'Parch'.

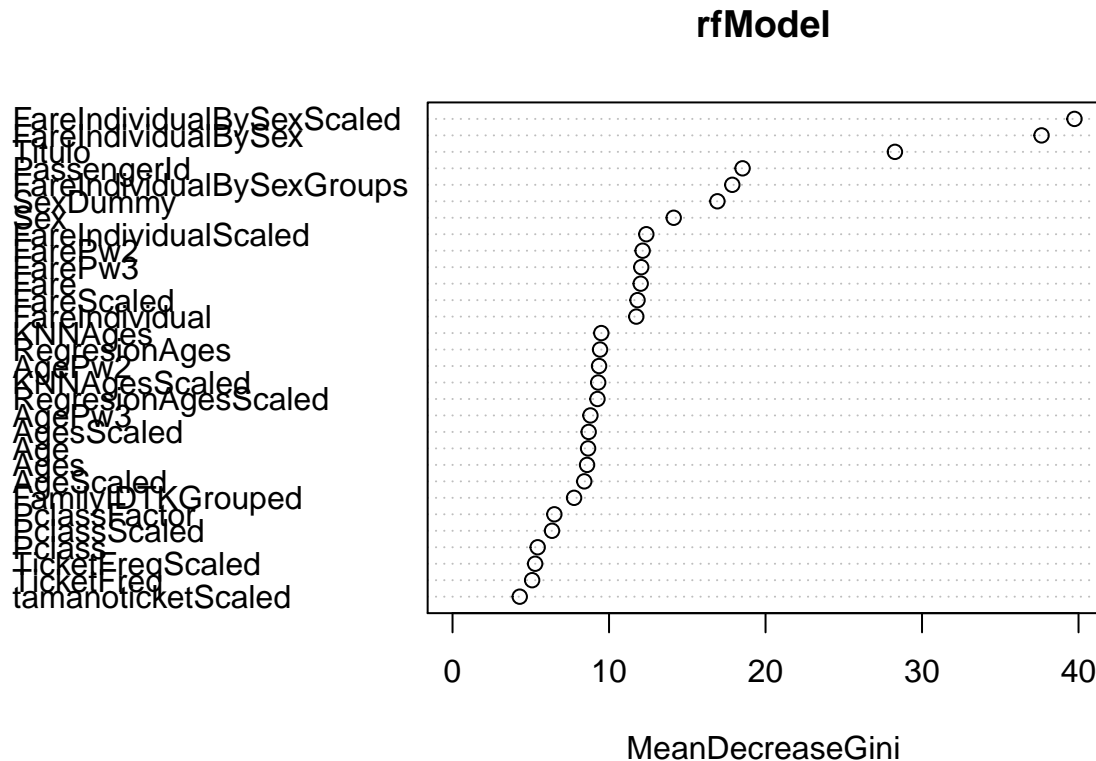
```
## [1] "Importancia de las variables"
```

```

##                               MeanDecreaseGini
## PassengerId                  18.5322089
## Pclass                       5.4413971
## Sex                          14.1321532
## Age                          8.6606190
## SibSp                        2.8868562
## Parch                        1.1570235
## Fare                         12.0225108
## Embarked                     2.8644493
## tamanoticket                 4.1010606
## TicketFreq                   5.0860676
## Titulo                       28.2684083
## FamSize                      3.8478243
## LoneWolfs                    0.7540797
## FamilyIDTKGrouped           7.7660895
## IsMother                     0.4410108
## FareGroups                   2.5226443
## FareIndividual               11.7421928
## FareIndGroups                3.5352225
## FareIndividualBySex          37.6356491
## FareIndividualBySexGroups    17.8752323
## ageGroupsByCut               2.0976972
## Ages                         8.5962136
## KNNAges                      9.5059961
## RegresionAges                9.4265706
## youngTravelAl                0.2369829
## PclassFactor                 6.5061408
## AgePw2                       9.3639177
## FarePw2                      12.1469618
## AgePw3                       8.8112974
## FarePw3                      12.0595040
## PclassScaled                 6.3584410
## AgeScaled                    8.4181092
## SibSpScaled                  2.6542732
## ParchScaled                  1.2044630
## FareScaled                    11.8217099
## tamanoticketScaled           4.2943865
## TicketFreqScaled             5.2798304
## FamSizeScaled                4.1272630
## FareIndividualScaled         12.3820225
## AgesScaled                   8.6994244
## KNNAgesScaled                9.3075847
## RegresionAgesScaled          9.2619201
## FareIndividualBySexScaled    39.7404963
## SexDummy                     16.9248790
## LoneWolfsDummy               1.2194964
## IsMotherDummy                0.4884667
## youngTravelAlDummy           0.3449758

## [1] "Importancia de las variables 2"

```



## 4.2. Comprobación de la normalidad y homogeneidad de la varianza.

No nos dió tiempo para completar este apartado.

## 4.3. Aplicación de pruebas estadísticas para comparar los grupos de datos. En función de los datos y el objetivo del estudio, aplicar pruebas de contraste de hipótesis, correlaciones, regresiones, etc. Aplicar al menos tres métodos de análisis diferentes.

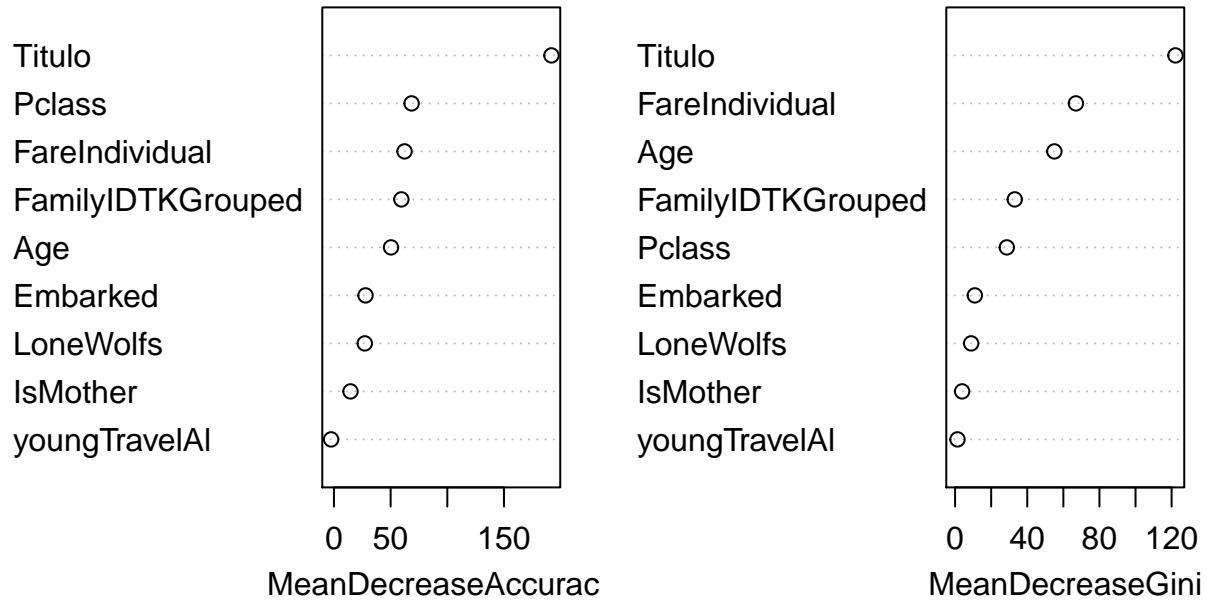
Vamos a utilizar diferentes métodos para predecir los resultados

### Predicción por Random Forest

Por este método lo aplicamos para predecir a partir de las variables: 'PassengerId', 'Survived', 'FamilyIDTKGrouped', 'Titulo', 'Fare', 'Age', 'Sex', 'Embarked', 'Port', 'Class', 'AgeScaled', 'SexScaled', 'EmbarkedScaled', 'PortScaled', 'ClassScaled', 'AgeGrouped', 'SexGrouped', 'EmbarkedGrouped', 'PortGrouped', 'ClassGrouped', 'AgeScaledGrouped', 'SexScaledGrouped', 'EmbarkedScaledGrouped', 'PortScaledGrouped', 'ClassScaledGrouped', 'AgeGroupedScaled', 'SexGroupedScaled', 'EmbarkedGroupedScaled', 'PortGroupedScaled', 'ClassGroupedScaled', 'AgeScaledGroupedScaled', 'SexScaledGroupedScaled', 'EmbarkedScaledGroupedScaled', 'PortScaledGroupedScaled', 'ClassScaledGroupedScaled'.

El resultado obtenido lo subimos a kaggle y nos dió: 0.779

fit



## Predicción por KNN

Este método obtiene los valores de Survived en función de sus vecinos más próximos que encuentra a partir de la distancia euclídea. Para usar este método escalamos todas las variables cuantitativas y pasamos todas las variables cualitativas a dummy, creando una variable numérica por cada uno de los factores de las variables.

El resultado obtenido por este método fue subido a kaggle y dio: 0.732

## Predicción por regresión logística

Por este método hemos obtenido en kaggle un 0.784

```
## [1] "Tabla de contingencia de sex vs. survived"

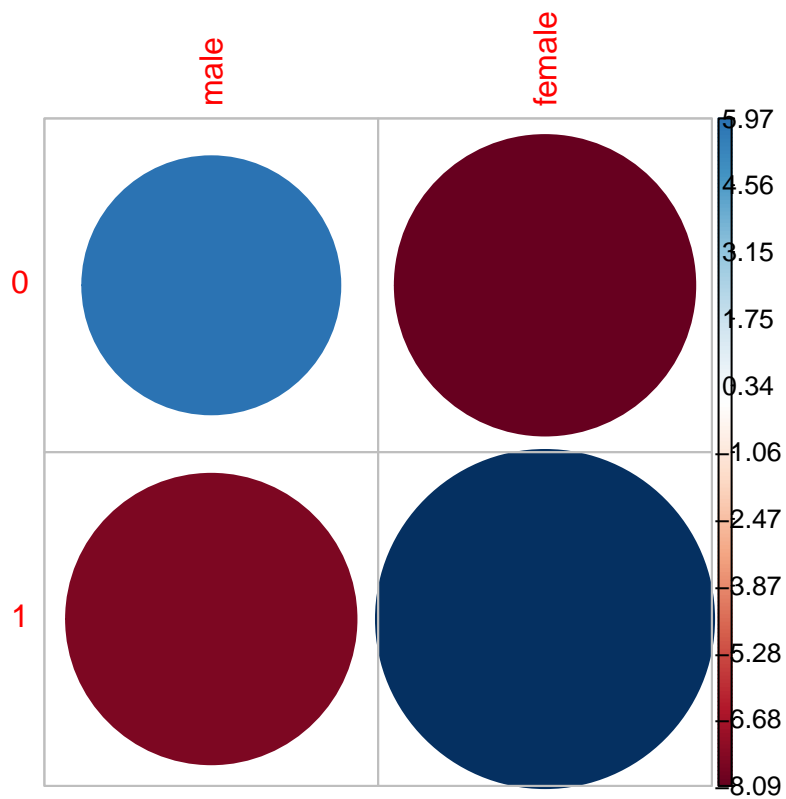
##           Sex
## Survived male female
##           0  468    81
##           1  109   233

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tabla_contingencia
## X-squared = 260.72, df = 1, p-value < 2.2e-16

## [1] "Odds ratio de sex vs. survived"

## [1] 12.35066
```





```
## [1] "Tabla de contingencia de Pclass vs. survived"
```

```
##      Pclass
```

```
## Survived  1  2  3
```

```
##      0  80  97 372
```

```
##      1 136  87 119
```

```
##
```

```
## Pearson's Chi-squared test
```

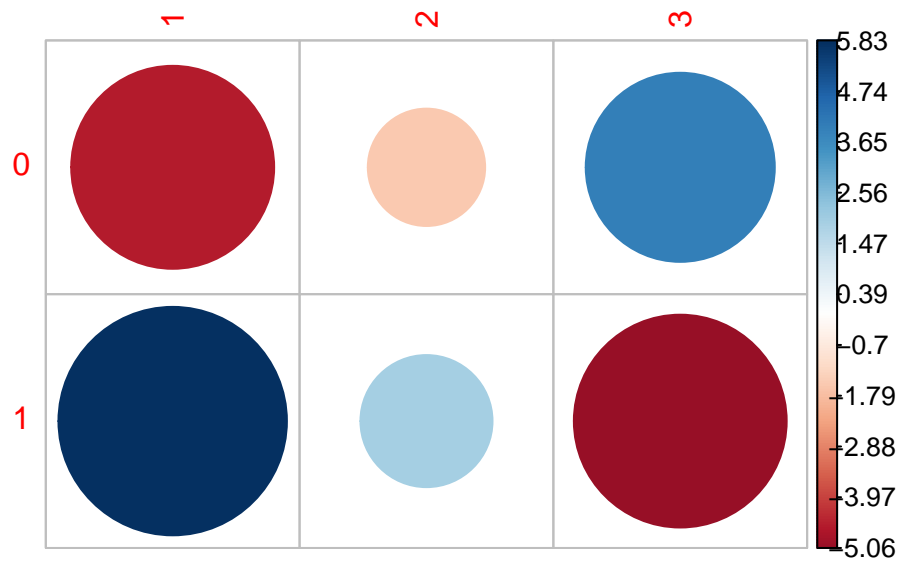
```
##
```

```
## data:  tabla_contingencia
```

```
## X-squared = 102.89, df = 2, p-value < 2.2e-16
```

```
## [1] "Odds ratio de sex Pclass survived"
```

```
## [1] 0.5275925
```



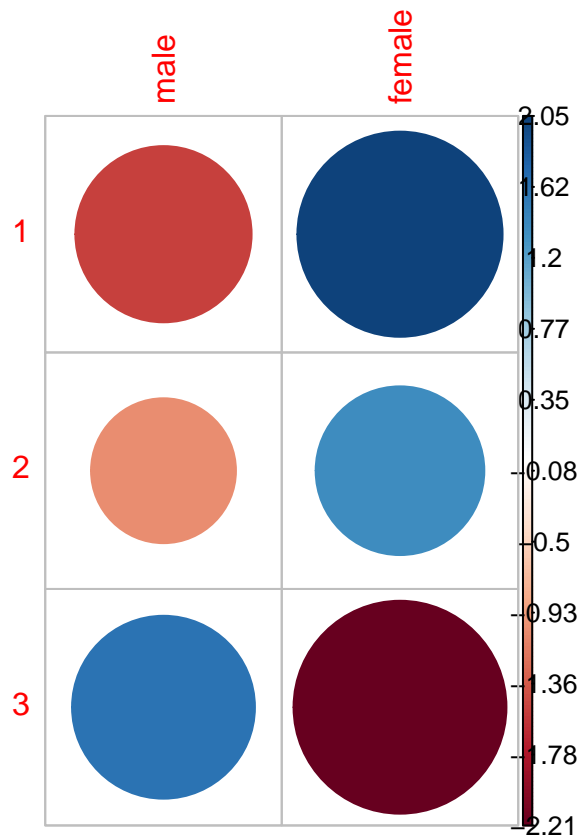
```
## [1] "Tabla de contingencia de sex vs. Pclass"

##      Sex
## Pclass male female
##      1  122    94
##      2  108    76
##      3  347   144

##
##  Pearson's Chi-squared test
##
## data:  tabla_contingencia
## X-squared = 16.971, df = 2, p-value = 0.0002064

## [1] "Odds ratio de sex vs. Pclass"

## [1] 0.9133176
```



```
##
## Call:
## glm(formula = Survived ~ FareIndividualBySexScaled + IsMotherDummy +
##     LoneWolfsDummy + TicketFreqScaled + ageGroupsByCut + Embarked +
##     PclassScaled, family = binomial, data = train2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4079  -0.5897  -0.3893   0.6484   2.8909
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.9606     0.3000   3.203 0.001362 **
## FareIndividualBySexScaled  1.3200     0.1110  11.892 < 2e-16 ***
## IsMotherDummy      0.6951     0.3888   1.788 0.073757 .
## LoneWolfsDummy    -0.2635     0.2438  -1.081 0.279773
## TicketFreqScaled  -0.4594     0.1391  -3.303 0.000955 ***
## ageGroupsByCutAG24 -1.5306     0.3289  -4.653 3.26e-06 ***
## ageGroupsByCutAG36 -1.9423     0.3511  -5.532 3.17e-08 ***
## ageGroupsByCutAG55 -2.3656     0.4647  -5.091 3.57e-07 ***
## ageGroupsByCutAG70 -3.7876     1.1512  -3.290 0.001002 **
## EmbarkedC         0.2802     0.2351   1.192 0.233346
## EmbarkedQ         0.4537     0.3260   1.392 0.164009
## PclassScaled      -0.6049     0.1070  -5.654 1.57e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1186.7  on 890  degrees of freedom
```

```
## Residual deviance: 787.9 on 879 degrees of freedom
## AIC: 811.9
##
## Number of Fisher Scoring iterations: 5
## [1] "Porcentaje de corte óptimo"
## [1] 0.5869116
## [1] "Error por fallo de mala clasificación corte= 0.58691162565729"
## [1] 0.193
## [1] "Error por fallo de mala clasificación corte 0.5"
## [1] 0.2009
## [1] "Tabla de contingencia de predicciones contra observaciones Survived"
##      as.factor(predicciones)
## Survived No Sobrevive Sobrevive
##      0      465      84
##      1      95      247
##
##      0      1
## 0.8469945 0.7222222
## [1] "Matriz de porcentaje de aciertos del modelo"
## [1] "acierto general del modelo"
## [1] 0.7991021
```

Table 2: Matriz de acierto del modelo. acierto global: 0.799102132435466

	<i>(Predicción peso)</i>		porcentaje_de_acierto
	No Sobrevive	Sobrevive	
0	465	84	0.8469945
1	95	247	0.7222222

```
## [1] "-----"
## [1] "Tabla de contingencia de predicciones optimizadas peso contra observaciones peso"
##      as.factor(prediccionesOptimizadas)
## Survived No Sobrevive Sobrevive
##      0      502      47
##      1      125      217
## [1] "-----"
## [1] "-----"
## [1] "Porcentaje de acierto general del modelo: 0.80695847362514"
## [1] "Matriz de porcentaje de aciertos del modelo optimizado"
```

Table 3: Matriz de acierto del modelo optimizado. acierto global: 0.80695847362514

	<i>(Predicción peso)</i>		porcentaje_de_acierto
	No Sobrevive	Sobrevive	
0	502	47	0.9143898
1	125	217	0.6345029

```
## [1] "variance inflation factors"
```

```

## FareIndividualBySexScaled      IsMotherDummy      LoneWolfsDummy
##           1.2141           1.3121           1.8684
##           TicketFreqScaled      ageGroupsByCutAG24      ageGroupsByCutAG36
##           1.9932           3.3705           3.4301
##           ageGroupsByCutAG55      ageGroupsByCutAG70      EmbarkedC
##           2.1091           1.1140           1.0717
##           EmbarkedQ      PclassScaled
##           1.1315           1.4733

## [1] "loglikelihood modelo"

## 'log Lik.' -393.9477 (df=12)

## Likelihood ratio test
##
## Model 1: Survived ~ FareIndividualBySexScaled + IsMotherDummy + LoneWolfsDummy +
##           TicketFreqScaled + ageGroupsByCut + Embarked + PclassScaled
## Model 2: Survived ~ 1
## #Df LogLik Df Chisq Pr(>Chisq)
## 1 12 -393.95
## 2 1 -593.33 -11 398.76 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] "loglikelihood Ratio test contra modelo con solo la constante"

## Likelihood ratio test
##
## Model 1: Survived ~ 1
## Model 2: Survived ~ FareIndividualBySexScaled + IsMotherDummy + LoneWolfsDummy +
##           TicketFreqScaled + ageGroupsByCut + Embarked + PclassScaled
## #Df LogLik Df Chisq Pr(>Chisq)
## 1 1 -593.33
## 2 12 -393.95 11 398.76 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] "MacFadden's Pseudo R^2"

## 'log Lik.' 0.3360367 (df=12)

## Wald test
##
## Model 1: Survived ~ 1
## Model 2: Survived ~ FareIndividualBySexScaled + IsMotherDummy + LoneWolfsDummy +
##           TicketFreqScaled + ageGroupsByCut + Embarked + PclassScaled
## Res.Df Df Chisq Pr(>Chisq)
## 1 890
## 2 879 11 238.86 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Call:
## concordance.formula(object = train2$Survived ~ predicciones)
##
## n= 891
## Concordance= 0.7846 se= 0.01434
## concordant discordant tied.x tied.y tied.xy
## 114855 7980 64923 62525 146212

## Overall accuracy = 0.799
##
## Confusion matrix

```

```

##          Predicted (cv)
## Actual    0      1
##          0 0.847 0.153
##          1 0.278 0.722

## [1] "Matriz de confusion a partir de los datos originales"

## $overall
## [1] 0.7991021
##
## $confusion
##          Predicted (cv)
## Actual    0      1
##          0 0.8469945 0.1530055
##          1 0.2777778 0.7222222
##
## $prior
##          0      1
## 0.6161616 0.3838384

## [1] "Matriz de confusion a partir de los valores del modelo"

## Overall accuracy = 0.003
##
## Confusion matrix
##          Predicted (cv)
## Actual  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
## [1,] 0.005 0.002 0.002 0.002 0.004 0.002 0.002 0.004 0.002 0.002 0.002 0.002
## [2,] 0.000 0.000 0.006 0.000 0.000 0.009 0.000 0.000 0.000 0.000 0.000 0.000
##          Predicted (cv)
## Actual  [,13] [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
## [1,] 0.002 0.002 0.002 0.005 0.002 0.002 0.002 0.002 0.004 0.002 0.002 0.002
## [2,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
##          Predicted (cv)
## Actual  [,25] [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34] [,35] [,36]
## [1,] 0.002 0.009 0.002 0.005 0.004 0.002 0.007 0.002 0.004 0.002 0.002 0.002
## [2,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
##          Predicted (cv)
## Actual  [,37] [,38] [,39] [,40] [,41] [,42] [,43] [,44] [,45] [,46] [,47] [,48]
## [1,] 0.002 0.004 0.002 0.031 0.002 0.002 0.002 0.035 0.000 0.004 0.002 0.005
## [2,] 0.000 0.003 0.000 0.000 0.000 0.000 0.006 0.003 0.003 0.000 0.000 0.000
##          Predicted (cv)
## Actual  [,49] [,50] [,51] [,52] [,53] [,54] [,55] [,56] [,57] [,58] [,59] [,60]
## [1,] 0.002 0.004 0.002 0.002 0.002 0.005 0.002 0.002 0.000 0.000 0.004 0.002
## [2,] 0.000 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.003 0.003 0.000 0.000
##          Predicted (cv)
## Actual  [,61] [,62] [,63] [,64] [,65] [,66] [,67] [,68] [,69] [,70] [,71] [,72]
## [1,] 0.002 0.002 0.013 0.009 0.002 0.002 0.007 0.002 0.002 0.002 0.002 0.007
## [2,] 0.000 0.000 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
##          Predicted (cv)
## Actual  [,73] [,74] [,75] [,76] [,77] [,78] [,79] [,80] [,81] [,82] [,83] [,84]
## [1,] 0.000 0.002 0.000 0.007 0.004 0.000 0.011 0.002 0.002 0.002 0.004 0.002
## [2,] 0.003 0.000 0.003 0.000 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.000
##          Predicted (cv)
## Actual  [,85] [,86] [,87] [,88] [,89] [,90] [,91] [,92] [,93] [,94] [,95] [,96]
## [1,] 0.011 0.007 0.002 0.005 0.009 0.002 0.005 0.002 0.027 0.002 0.009 0.002
## [2,] 0.000 0.006 0.000 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.006 0.000
##          Predicted (cv)
## Actual  [,97] [,98] [,99] [,100] [,101] [,102] [,103] [,104] [,105] [,106]
## [1,] 0.024 0.004 0.002 0.007 0.002 0.002 0.002 0.009 0.002 0.002
## [2,] 0.009 0.000 0.000 0.000 0.000 0.000 0.000 0.003 0.000 0.000

```

```

## Predicted (cv)
## Actual [,107] [,108] [,109] [,110] [,111] [,112] [,113] [,114] [,115] [,116]
## [1,] 0.002 0.002 0.018 0.002 0.007 0.002 0.002 0.004 0.002 0.002
## [2,] 0.000 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## Predicted (cv)
## Actual [,117] [,118] [,119] [,120] [,121] [,122] [,123] [,124] [,125] [,126]
## [1,] 0.007 0.000 0.002 0.002 0.002 0.002 0.002 0.002 0.004 0.002
## [2,] 0.006 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## Predicted (cv)
## Actual [,127] [,128] [,129] [,130] [,131] [,132] [,133] [,134] [,135] [,136]
## [1,] 0.002 0.004 0.002 0.002 0.009 0.002 0.002 0.002 0.002 0.004
## [2,] 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.003
## Predicted (cv)
## Actual [,137] [,138] [,139] [,140] [,141] [,142] [,143] [,144] [,145] [,146]
## [1,] 0.007 0.002 0.004 0.002 0.004 0.002 0.002 0.000 0.002 0.002
## [2,] 0.003 0.003 0.000 0.000 0.000 0.000 0.003 0.003 0.000 0.000
## Predicted (cv)
## Actual [,147] [,148] [,149] [,150] [,151] [,152] [,153] [,154] [,155] [,156]
## [1,] 0.002 0.004 0.000 0.004 0.002 0.002 0.004 0.004 0.002 0.002
## [2,] 0.000 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## Predicted (cv)
## Actual [,157] [,158] [,159] [,160] [,161] [,162] [,163] [,164] [,165] [,166]
## [1,] 0.004 0.002 0.002 0.011 0.002 0.002 0.011 0.002 0.002 0.002
## [2,] 0.003 0.000 0.000 0.009 0.000 0.000 0.000 0.000 0.000 0.000
## Predicted (cv)
## Actual [,167] [,168] [,169] [,170] [,171] [,172] [,173] [,174] [,175] [,176]
## [1,] 0.002 0.002 0.002 0.000 0.002 0.007 0.002 0.002 0.002 0.000
## [2,] 0.000 0.000 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.003
## Predicted (cv)
## Actual [,177] [,178] [,179] [,180] [,181] [,182] [,183] [,184] [,185] [,186]
## [1,] 0.016 0.000 0.002 0.007 0.004 0.002 0.002 0.002 0.000 0.002
## [2,] 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.003 0.000
## Predicted (cv)
## Actual [,187] [,188] [,189] [,190] [,191] [,192] [,193] [,194] [,195] [,196]
## [1,] 0.027 0.002 0.004 0.002 0.002 0.002 0.002 0.000 0.002 0.004
## [2,] 0.003 0.000 0.003 0.000 0.000 0.000 0.000 0.003 0.000 0.000
## Predicted (cv)
## Actual [,197] [,198] [,199] [,200] [,201] [,202] [,203] [,204] [,205] [,206]
## [1,] 0.002 0.002 0.000 0.000 0.002 0.000 0.002 0.000 0.002 0.002
## [2,] 0.000 0.000 0.006 0.003 0.000 0.003 0.000 0.003 0.000 0.000
## Predicted (cv)
## Actual [,207] [,208] [,209] [,210] [,211] [,212] [,213] [,214] [,215] [,216]
## [1,] 0.002 0.002 0.002 0.007 0.000 0.002 0.002 0.000 0.000 0.005
## [2,] 0.000 0.000 0.000 0.003 0.003 0.000 0.000 0.006 0.003 0.000
## Predicted (cv)
## Actual [,217] [,218] [,219] [,220] [,221] [,222] [,223] [,224] [,225] [,226]
## [1,] 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002
## [2,] 0.003 0.000 0.000 0.000 0.003 0.000 0.000 0.000 0.000 0.000
## Predicted (cv)
## Actual [,227] [,228] [,229] [,230] [,231] [,232] [,233] [,234] [,235] [,236]
## [1,] 0.000 0.002 0.000 0.000 0.000 0.000 0.002 0.002 0.002 0.002
## [2,] 0.003 0.000 0.003 0.003 0.003 0.003 0.000 0.000 0.000 0.000
## Predicted (cv)
## Actual [,237] [,238] [,239] [,240] [,241] [,242] [,243] [,244] [,245] [,246]
## [1,] 0.002 0.005 0.000 0.000 0.000 0.004 0.002 0.002 0.000 0.002
## [2,] 0.000 0.003 0.003 0.006 0.006 0.012 0.000 0.000 0.003 0.000
## Predicted (cv)
## Actual [,247] [,248] [,249] [,250] [,251] [,252] [,253] [,254] [,255] [,256]
## [1,] 0.002 0.000 0.002 0.002 0.002 0.002 0.002 0.002 0.000 0.000

```

```

## [2,] 0.000 0.003 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.003 0.003
## Predicted (cv)
## Actual [,257] [,258] [,259] [,260] [,261] [,262] [,263] [,264] [,265] [,266]
## [1,] 0.000 0.002 0.002 0.002 0.002 0.002 0.000 0.000 0.000 0.000
## [2,] 0.006 0.000 0.000 0.000 0.000 0.000 0.003 0.006 0.003 0.003
## Predicted (cv)
## Actual [,267] [,268] [,269] [,270] [,271] [,272] [,273] [,274] [,275] [,276]
## [1,] 0.000 0.000 0.002 0.002 0.000 0.002 0.000 0.002 0.000 0.002
## [2,] 0.003 0.006 0.000 0.000 0.003 0.000 0.003 0.000 0.003 0.000
## Predicted (cv)
## Actual [,277] [,278] [,279] [,280] [,281] [,282] [,283] [,284] [,285] [,286]
## [1,] 0.005 0.002 0.002 0.002 0.000 0.002 0.002 0.002 0.002 0.000
## [2,] 0.000 0.000 0.000 0.003 0.003 0.000 0.000 0.000 0.000 0.003
## Predicted (cv)
## Actual [,287] [,288] [,289] [,290] [,291] [,292] [,293] [,294] [,295] [,296]
## [1,] 0.000 0.000 0.002 0.002 0.000 0.000 0.002 0.000 0.000 0.002
## [2,] 0.003 0.003 0.003 0.000 0.003 0.003 0.000 0.003 0.003 0.000
## Predicted (cv)
## Actual [,297] [,298] [,299] [,300] [,301] [,302] [,303] [,304] [,305] [,306]
## [1,] 0.002 0.002 0.002 0.002 0.000 0.002 0.000 0.000 0.000 0.000
## [2,] 0.000 0.000 0.000 0.000 0.003 0.000 0.003 0.003 0.003 0.003
## Predicted (cv)
## Actual [,307] [,308] [,309] [,310] [,311] [,312] [,313] [,314] [,315] [,316]
## [1,] 0.002 0.002 0.000 0.000 0.000 0.004 0.002 0.002 0.002 0.002
## [2,] 0.000 0.003 0.003 0.006 0.003 0.003 0.003 0.000 0.000 0.006
## Predicted (cv)
## Actual [,317] [,318] [,319] [,320] [,321] [,322] [,323] [,324] [,325] [,326]
## [1,] 0.002 0.000 0.000 0.002 0.002 0.004 0.000 0.002 0.000 0.002
## [2,] 0.000 0.003 0.003 0.000 0.000 0.000 0.003 0.000 0.003 0.000
## Predicted (cv)
## Actual [,327] [,328] [,329] [,330] [,331] [,332] [,333] [,334] [,335] [,336]
## [1,] 0.004 0.002 0.000 0.002 0.002 0.000 0.002 0.002 0.002 0.007
## [2,] 0.000 0.000 0.003 0.000 0.000 0.003 0.000 0.000 0.000 0.003
## Predicted (cv)
## Actual [,337] [,338] [,339] [,340] [,341] [,342] [,343] [,344] [,345] [,346]
## [1,] 0.002 0.000 0.002 0.000 0.000 0.000 0.002 0.002 0.000 0.002
## [2,] 0.000 0.003 0.000 0.003 0.003 0.003 0.000 0.000 0.003 0.000
## Predicted (cv)
## Actual [,347] [,348] [,349] [,350] [,351] [,352] [,353] [,354] [,355] [,356]
## [1,] 0.000 0.000 0.002 0.002 0.000 0.000 0.000 0.004 0.002 0.000
## [2,] 0.003 0.003 0.000 0.000 0.003 0.003 0.003 0.000 0.000 0.003
## Predicted (cv)
## Actual [,357] [,358] [,359] [,360] [,361] [,362] [,363] [,364] [,365] [,366]
## [1,] 0.002 0.000 0.002 0.000 0.000 0.000 0.002 0.002 0.004 0.000
## [2,] 0.003 0.003 0.000 0.003 0.003 0.003 0.000 0.000 0.000 0.003
## Predicted (cv)
## Actual [,367] [,368] [,369] [,370] [,371] [,372] [,373] [,374] [,375] [,376]
## [1,] 0.000 0.004 0.000 0.002 0.000 0.000 0.000 0.002 0.000 0.000
## [2,] 0.003 0.000 0.003 0.000 0.003 0.003 0.003 0.000 0.003 0.003
## Predicted (cv)
## Actual [,377] [,378] [,379] [,380] [,381] [,382] [,383] [,384] [,385] [,386]
## [1,] 0.005 0.002 0.000 0.000 0.000 0.000 0.000 0.002 0.002 0.002
## [2,] 0.023 0.000 0.003 0.003 0.012 0.006 0.003 0.000 0.018 0.006
## Predicted (cv)
## Actual [,387] [,388] [,389] [,390] [,391] [,392] [,393] [,394] [,395] [,396]
## [1,] 0.000 0.002 0.000 0.000 0.000 0.002 0.002 0.000 0.004 0.000
## [2,] 0.003 0.000 0.003 0.003 0.003 0.000 0.000 0.003 0.000 0.003
## Predicted (cv)
## Actual [,397] [,398] [,399] [,400] [,401] [,402] [,403] [,404] [,405] [,406]

```



```

## [1,] 0.002 0.002 0.002 0.000 0.002 0.000 0.000 0.000 0.000 0.000 0.002
## [2,] 0.000 0.000 0.000 0.003 0.000 0.003 0.003 0.003 0.003 0.012 0.003
## Predicted (cv)
## Actual [,407] [,408] [,409] [,410] [,411] [,412] [,413] [,414] [,415] [,416]
## [1,] 0.000 0.000 0.000 0.000 0.002 0.000 0.000 0.000 0.000 0.000
## [2,] 0.003 0.003 0.003 0.003 0.015 0.018 0.003 0.003 0.003 0.006
## Predicted (cv)
## Actual [,417] [,418] [,419] [,420] [,421] [,422] [,423] [,424] [,425] [,426]
## [1,] 0.000 0.000 0.000 0.000 0.000 0.002 0.000 0.000 0.000 0.000
## [2,] 0.003 0.003 0.003 0.003 0.003 0.000 0.003 0.003 0.003 0.009
## Predicted (cv)
## Actual [,427] [,428] [,429] [,430] [,431] [,432] [,433] [,434] [,435] [,436]
## [1,] 0.000 0.000 0.000 0.000 0.000 0.000 0.002 0.000 0.000 0.000
## [2,] 0.003 0.003 0.003 0.003 0.003 0.003 0.000 0.003 0.003 0.003
## Predicted (cv)
## Actual [,437] [,438] [,439] [,440] [,441] [,442] [,443] [,444] [,445] [,446]
## [1,] 0.000 0.000 0.000 0.000 0.000 0.000 0.002 0.002 0.000 0.000
## [2,] 0.003 0.003 0.003 0.006 0.003 0.003 0.000 0.000 0.003 0.003
## Predicted (cv)
## Actual [,447] [,448] [,449] [,450] [,451] [,452] [,453] [,454] [,455] [,456]
## [1,] 0.000 0.000 0.002 0.000 0.000 0.000 0.000 0.002 0.000 0.000
## [2,] 0.003 0.003 0.000 0.003 0.003 0.003 0.006 0.003 0.003 0.003
## Predicted (cv)
## Actual [,457] [,458] [,459] [,460] [,461] [,462] [,463] [,464] [,465] [,466]
## [1,] 0.000 0.002 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [2,] 0.003 0.000 0.003 0.003 0.003 0.006 0.003 0.003 0.006 0.003
## Predicted (cv)
## Actual [,467] [,468] [,469] [,470] [,471] [,472] [,473] [,474] [,475] [,476]
## [1,] 0.000 0.002 0.002 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [2,] 0.003 0.003 0.000 0.003 0.003 0.003 0.006 0.003 0.003 0.003
## Predicted (cv)
## Actual [,477] [,478] [,479] [,480] [,481] [,482] [,483] [,484] [,485] [,486]
## [1,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [2,] 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003
## Predicted (cv)
## Actual [,487] [,488] [,489] [,490] [,491] [,492] [,493] [,494] [,495] [,496]
## [1,] 0.000 0.000 0.000 0.002 0.000 0.000 0.000 0.000 0.000 0.000
## [2,] 0.003 0.006 0.006 0.000 0.003 0.003 0.003 0.003 0.003 0.003
## Predicted (cv)
## Actual [,497] [,498] [,499] [,500] [,501] [,502] [,503] [,504] [,505] [,506]
## [1,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [2,] 0.003 0.003 0.006 0.003 0.003 0.003 0.003 0.003 0.003 0.003
## Predicted (cv)
## Actual [,507] [,508] [,509] [,510] [,511] [,512] [,513] [,514] [,515] [,516]
## [1,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [2,] 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003
## Predicted (cv)
## Actual [,517] [,518] [,519] [,520] [,521] [,522] [,523] [,524] [,525] [,526]
## [1,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.002 0.000 0.000
## [2,] 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.000 0.006 0.003
## Predicted (cv)
## Actual [,527] [,528] [,529] [,530] [,531] [,532] [,533] [,534] [,535] [,536]
## [1,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [2,] 0.003 0.003 0.003 0.006 0.003 0.003 0.003 0.003 0.003 0.006
## Predicted (cv)
## Actual [,537] [,538] [,539] [,540] [,541] [,542] [,543] [,544] [,545] [,546]
## [1,] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [2,] 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003
## Predicted (cv)

```

```

## Actual [,547]
## [1,] 0.000
## [2,] 0.003

## $overall
## [1] 0.003367003
##
## $confusion
## Predicted (cv)
## Actual [,1] [,2] [,3] [,4] [,5] [,6]
## [1,] 0.005464481 0.001821494 0.001821494 0.001821494 0.003642987 0.001821494
## [2,] 0.000000000 0.000000000 0.005847953 0.000000000 0.000000000 0.008771930
## Predicted (cv)
## Actual [,7] [,8] [,9] [,10] [,11] [,12]
## [1,] 0.001821494 0.003642987 0.001821494 0.001821494 0.001821494 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [,13] [,14] [,15] [,16] [,17] [,18]
## [1,] 0.001821494 0.001821494 0.001821494 0.005464481 0.001821494 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [,19] [,20] [,21] [,22] [,23] [,24]
## [1,] 0.001821494 0.001821494 0.003642987 0.001821494 0.001821494 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [,25] [,26] [,27] [,28] [,29] [,30]
## [1,] 0.001821494 0.009107468 0.001821494 0.005464481 0.003642987 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [,31] [,32] [,33] [,34] [,35] [,36]
## [1,] 0.007285974 0.001821494 0.003642987 0.001821494 0.001821494 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [,37] [,38] [,39] [,40] [,41] [,42]
## [1,] 0.001821494 0.003642987 0.001821494 0.03096539 0.001821494 0.001821494
## [2,] 0.000000000 0.002923977 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [,43] [,44] [,45] [,46] [,47] [,48]
## [1,] 0.001821494 0.034608379 0.000000000 0.003642987 0.001821494 0.005464481
## [2,] 0.005847953 0.002923977 0.002923977 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [,49] [,50] [,51] [,52] [,53] [,54]
## [1,] 0.001821494 0.003642987 0.001821494 0.001821494 0.001821494 0.005464481
## [2,] 0.000000000 0.000000000 0.002923977 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [,55] [,56] [,57] [,58] [,59] [,60]
## [1,] 0.001821494 0.001821494 0.000000000 0.000000000 0.003642987 0.001821494
## [2,] 0.000000000 0.000000000 0.002923977 0.002923977 0.000000000 0.000000000
## Predicted (cv)
## Actual [,61] [,62] [,63] [,64] [,65] [,66]
## [1,] 0.001821494 0.001821494 0.01275046 0.009107468 0.001821494 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.002923977 0.000000000 0.000000000
## Predicted (cv)
## Actual [,67] [,68] [,69] [,70] [,71] [,72]
## [1,] 0.007285974 0.001821494 0.001821494 0.001821494 0.001821494 0.007285974
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [,73] [,74] [,75] [,76] [,77] [,78]
## [1,] 0.000000000 0.001821494 0.000000000 0.007285974 0.003642987 0.000000000

```

```

## [2,] 0.002923977 0.000000000 0.002923977 0.000000000 0.000000000 0.002923977
## Predicted (cv)
## Actual [ ,79] [ ,80] [ ,81] [ ,82] [ ,83] [ ,84]
## [1,] 0.01092896 0.001821494 0.001821494 0.001821494 0.003642987 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [ ,85] [ ,86] [ ,87] [ ,88] [ ,89] [ ,90]
## [1,] 0.01092896 0.007285974 0.001821494 0.005464481 0.009107468 0.001821494
## [2,] 0.000000000 0.005847953 0.000000000 0.000000000 0.002923977 0.000000000
## Predicted (cv)
## Actual [ ,91] [ ,92] [ ,93] [ ,94] [ ,95] [ ,96]
## [1,] 0.005464481 0.001821494 0.0273224 0.001821494 0.009107468 0.001821494
## [2,] 0.000000000 0.000000000 0.00000000 0.000000000 0.005847953 0.000000000
## Predicted (cv)
## Actual [ ,97] [ ,98] [ ,99] [ ,100] [ ,101] [ ,102]
## [1,] 0.02367942 0.003642987 0.001821494 0.007285974 0.001821494 0.001821494
## [2,] 0.00877193 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [ ,103] [ ,104] [ ,105] [ ,106] [ ,107] [ ,108]
## [1,] 0.001821494 0.009107468 0.001821494 0.001821494 0.001821494 0.001821494
## [2,] 0.000000000 0.002923977 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [ ,109] [ ,110] [ ,111] [ ,112] [ ,113] [ ,114]
## [1,] 0.0018214936 0.001821494 0.007285974 0.001821494 0.001821494 0.003642987
## [2,] 0.002923977 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [ ,115] [ ,116] [ ,117] [ ,118] [ ,119] [ ,120]
## [1,] 0.001821494 0.001821494 0.007285974 0.000000000 0.001821494 0.001821494
## [2,] 0.000000000 0.000000000 0.005847953 0.002923977 0.000000000 0.000000000
## Predicted (cv)
## Actual [ ,121] [ ,122] [ ,123] [ ,124] [ ,125] [ ,126]
## [1,] 0.001821494 0.001821494 0.001821494 0.001821494 0.003642987 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [ ,127] [ ,128] [ ,129] [ ,130] [ ,131] [ ,132]
## [1,] 0.001821494 0.003642987 0.001821494 0.001821494 0.009107468 0.001821494
## [2,] 0.002923977 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [ ,133] [ ,134] [ ,135] [ ,136] [ ,137] [ ,138]
## [1,] 0.001821494 0.001821494 0.001821494 0.003642987 0.007285974 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,139] [ ,140] [ ,141] [ ,142] [ ,143] [ ,144]
## [1,] 0.003642987 0.001821494 0.003642987 0.001821494 0.001821494 0.000000000
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,145] [ ,146] [ ,147] [ ,148] [ ,149] [ ,150]
## [1,] 0.001821494 0.001821494 0.001821494 0.003642987 0.000000000 0.003642987
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.002923977 0.000000000
## Predicted (cv)
## Actual [ ,151] [ ,152] [ ,153] [ ,154] [ ,155] [ ,156]
## [1,] 0.001821494 0.001821494 0.003642987 0.003642987 0.001821494 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [ ,157] [ ,158] [ ,159] [ ,160] [ ,161] [ ,162]
## [1,] 0.003642987 0.001821494 0.001821494 0.01092896 0.001821494 0.001821494
## [2,] 0.002923977 0.000000000 0.000000000 0.00877193 0.000000000 0.000000000
## Predicted (cv)
## Actual [ ,163] [ ,164] [ ,165] [ ,166] [ ,167] [ ,168]

```

```

## [1,] 0.01092896 0.001821494 0.001821494 0.001821494 0.001821494 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [1,169] [1,170] [1,171] [1,172] [1,173] [1,174]
## [1,] 0.001821494 0.000000000 0.001821494 0.007285974 0.001821494 0.001821494
## [2,] 0.000000000 0.002923977 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [1,175] [1,176] [1,177] [1,178] [1,179] [1,180]
## [1,] 0.001821494 0.000000000 0.01639344 0.000000000 0.001821494 0.007285974
## [2,] 0.000000000 0.002923977 0.000000000 0.002923977 0.000000000 0.000000000
## Predicted (cv)
## Actual [1,181] [1,182] [1,183] [1,184] [1,185] [1,186]
## [1,] 0.003642987 0.001821494 0.001821494 0.001821494 0.000000000 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.002923977 0.000000000
## Predicted (cv)
## Actual [1,187] [1,188] [1,189] [1,190] [1,191] [1,192]
## [1,] 0.027322404 0.001821494 0.003642987 0.001821494 0.001821494 0.001821494
## [2,] 0.002923977 0.000000000 0.002923977 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [1,193] [1,194] [1,195] [1,196] [1,197] [1,198]
## [1,] 0.001821494 0.000000000 0.001821494 0.003642987 0.001821494 0.001821494
## [2,] 0.000000000 0.002923977 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual [1,199] [1,200] [1,201] [1,202] [1,203] [1,204]
## [1,] 0.000000000 0.000000000 0.001821494 0.000000000 0.001821494 0.000000000
## [2,] 0.005847953 0.002923977 0.000000000 0.002923977 0.000000000 0.002923977
## Predicted (cv)
## Actual [1,205] [1,206] [1,207] [1,208] [1,209] [1,210]
## [1,] 0.001821494 0.001821494 0.001821494 0.001821494 0.001821494 0.007285974
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.002923977
## Predicted (cv)
## Actual [1,211] [1,212] [1,213] [1,214] [1,215] [1,216]
## [1,] 0.000000000 0.001821494 0.001821494 0.000000000 0.000000000 0.005464481
## [2,] 0.002923977 0.000000000 0.000000000 0.005847953 0.002923977 0.000000000
## Predicted (cv)
## Actual [1,217] [1,218] [1,219] [1,220] [1,221] [1,222]
## [1,] 0.001821494 0.001821494 0.001821494 0.001821494 0.001821494 0.001821494
## [2,] 0.002923977 0.000000000 0.000000000 0.000000000 0.002923977 0.000000000
## Predicted (cv)
## Actual [1,223] [1,224] [1,225] [1,226] [1,227] [1,228]
## [1,] 0.001821494 0.001821494 0.001821494 0.001821494 0.000000000 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.002923977 0.000000000
## Predicted (cv)
## Actual [1,229] [1,230] [1,231] [1,232] [1,233] [1,234]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.001821494 0.001821494
## [2,] 0.002923977 0.002923977 0.002923977 0.002923977 0.000000000 0.000000000
## Predicted (cv)
## Actual [1,235] [1,236] [1,237] [1,238] [1,239] [1,240]
## [1,] 0.001821494 0.001821494 0.001821494 0.005464481 0.000000000 0.000000000
## [2,] 0.000000000 0.000000000 0.000000000 0.002923977 0.002923977 0.005847953
## Predicted (cv)
## Actual [1,241] [1,242] [1,243] [1,244] [1,245] [1,246]
## [1,] 0.000000000 0.003642987 0.001821494 0.001821494 0.000000000 0.001821494
## [2,] 0.005847953 0.011695906 0.000000000 0.000000000 0.002923977 0.000000000
## Predicted (cv)
## Actual [1,247] [1,248] [1,249] [1,250] [1,251] [1,252]
## [1,] 0.001821494 0.000000000 0.001821494 0.001821494 0.001821494 0.001821494
## [2,] 0.000000000 0.002923977 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)

```

```

## Actual      [,253]      [,254]      [,255]      [,256]      [,257]      [,258]
## [1,] 0.001821494 0.001821494 0.000000000 0.000000000 0.000000000 0.001821494
## [2,] 0.000000000 0.000000000 0.002923977 0.002923977 0.005847953 0.000000000
## Predicted (cv)
## Actual      [,259]      [,260]      [,261]      [,262]      [,263]      [,264]
## [1,] 0.001821494 0.001821494 0.001821494 0.001821494 0.000000000 0.000000000
## [2,] 0.000000000 0.000000000 0.000000000 0.000000000 0.002923977 0.005847953
## Predicted (cv)
## Actual      [,265]      [,266]      [,267]      [,268]      [,269]      [,270]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.001821494 0.001821494
## [2,] 0.002923977 0.002923977 0.002923977 0.005847953 0.000000000 0.000000000
## Predicted (cv)
## Actual      [,271]      [,272]      [,273]      [,274]      [,275]      [,276]
## [1,] 0.000000000 0.001821494 0.000000000 0.001821494 0.000000000 0.001821494
## [2,] 0.002923977 0.000000000 0.002923977 0.000000000 0.002923977 0.000000000
## Predicted (cv)
## Actual      [,277]      [,278]      [,279]      [,280]      [,281]      [,282]
## [1,] 0.005464481 0.001821494 0.001821494 0.001821494 0.000000000 0.001821494
## [2,] 0.000000000 0.000000000 0.000000000 0.002923977 0.002923977 0.000000000
## Predicted (cv)
## Actual      [,283]      [,284]      [,285]      [,286]      [,287]      [,288]
## [1,] 0.001821494 0.001821494 0.001821494 0.000000000 0.000000000 0.000000000
## [2,] 0.000000000 0.000000000 0.000000000 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual      [,289]      [,290]      [,291]      [,292]      [,293]      [,294]
## [1,] 0.001821494 0.001821494 0.000000000 0.000000000 0.001821494 0.000000000
## [2,] 0.002923977 0.000000000 0.002923977 0.002923977 0.000000000 0.002923977
## Predicted (cv)
## Actual      [,295]      [,296]      [,297]      [,298]      [,299]      [,300]
## [1,] 0.000000000 0.001821494 0.001821494 0.001821494 0.001821494 0.001821494
## [2,] 0.002923977 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## Predicted (cv)
## Actual      [,301]      [,302]      [,303]      [,304]      [,305]      [,306]
## [1,] 0.000000000 0.001821494 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.000000000 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual      [,307]      [,308]      [,309]      [,310]      [,311]      [,312]
## [1,] 0.001821494 0.001821494 0.000000000 0.000000000 0.000000000 0.003642987
## [2,] 0.000000000 0.002923977 0.002923977 0.005847953 0.002923977 0.002923977
## Predicted (cv)
## Actual      [,313]      [,314]      [,315]      [,316]      [,317]      [,318]
## [1,] 0.001821494 0.001821494 0.001821494 0.001821494 0.001821494 0.000000000
## [2,] 0.002923977 0.000000000 0.000000000 0.005847953 0.000000000 0.002923977
## Predicted (cv)
## Actual      [,319]      [,320]      [,321]      [,322]      [,323]      [,324]
## [1,] 0.000000000 0.001821494 0.001821494 0.003642987 0.000000000 0.001821494
## [2,] 0.002923977 0.000000000 0.000000000 0.000000000 0.002923977 0.000000000
## Predicted (cv)
## Actual      [,325]      [,326]      [,327]      [,328]      [,329]      [,330]
## [1,] 0.000000000 0.001821494 0.003642987 0.001821494 0.000000000 0.001821494
## [2,] 0.002923977 0.000000000 0.000000000 0.000000000 0.002923977 0.000000000
## Predicted (cv)
## Actual      [,331]      [,332]      [,333]      [,334]      [,335]      [,336]
## [1,] 0.001821494 0.000000000 0.001821494 0.001821494 0.007285974 0.000000000
## [2,] 0.000000000 0.002923977 0.000000000 0.000000000 0.000000000 0.002923977
## Predicted (cv)
## Actual      [,337]      [,338]      [,339]      [,340]      [,341]      [,342]
## [1,] 0.001821494 0.000000000 0.001821494 0.000000000 0.000000000 0.000000000
## [2,] 0.000000000 0.002923977 0.000000000 0.002923977 0.002923977 0.002923977

```

```

## Predicted (cv)
## Actual      [,343]      [,344]      [,345]      [,346]      [,347]      [,348]
## [1,] 0.001821494 0.001821494 0.000000000 0.001821494 0.000000000 0.000000000
## [2,] 0.000000000 0.000000000 0.002923977 0.000000000 0.002923977 0.002923977
## Predicted (cv)
## Actual      [,349]      [,350]      [,351]      [,352]      [,353]      [,354]
## [1,] 0.001821494 0.001821494 0.000000000 0.000000000 0.000000000 0.003642987
## [2,] 0.000000000 0.000000000 0.002923977 0.002923977 0.002923977 0.000000000
## Predicted (cv)
## Actual      [,355]      [,356]      [,357]      [,358]      [,359]      [,360]
## [1,] 0.001821494 0.000000000 0.001821494 0.000000000 0.001821494 0.000000000
## [2,] 0.000000000 0.002923977 0.002923977 0.002923977 0.000000000 0.002923977
## Predicted (cv)
## Actual      [,361]      [,362]      [,363]      [,364]      [,365]      [,366]
## [1,] 0.000000000 0.000000000 0.001821494 0.001821494 0.003642987 0.000000000
## [2,] 0.002923977 0.002923977 0.000000000 0.000000000 0.000000000 0.002923977
## Predicted (cv)
## Actual      [,367]      [,368]      [,369]      [,370]      [,371]      [,372]
## [1,] 0.000000000 0.003642987 0.000000000 0.001821494 0.000000000 0.000000000
## [2,] 0.002923977 0.000000000 0.002923977 0.000000000 0.002923977 0.002923977
## Predicted (cv)
## Actual      [,373]      [,374]      [,375]      [,376]      [,377]      [,378]
## [1,] 0.000000000 0.001821494 0.000000000 0.000000000 0.005464481 0.001821494
## [2,] 0.002923977 0.000000000 0.002923977 0.002923977 0.023391813 0.000000000
## Predicted (cv)
## Actual      [,379]      [,380]      [,381]      [,382]      [,383]      [,384]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.001821494
## [2,] 0.002923977 0.002923977 0.01169591 0.005847953 0.002923977 0.000000000
## Predicted (cv)
## Actual      [,385]      [,386]      [,387]      [,388]      [,389]      [,390]
## [1,] 0.001821494 0.001821494 0.000000000 0.001821494 0.000000000 0.000000000
## [2,] 0.017543860 0.005847953 0.002923977 0.000000000 0.002923977 0.002923977
## Predicted (cv)
## Actual      [,391]      [,392]      [,393]      [,394]      [,395]      [,396]
## [1,] 0.000000000 0.001821494 0.001821494 0.000000000 0.003642987 0.000000000
## [2,] 0.002923977 0.000000000 0.000000000 0.002923977 0.000000000 0.002923977
## Predicted (cv)
## Actual      [,397]      [,398]      [,399]      [,400]      [,401]      [,402]
## [1,] 0.001821494 0.001821494 0.001821494 0.000000000 0.001821494 0.000000000
## [2,] 0.000000000 0.000000000 0.000000000 0.002923977 0.000000000 0.002923977
## Predicted (cv)
## Actual      [,403]      [,404]      [,405]      [,406]      [,407]      [,408]
## [1,] 0.000000000 0.000000000 0.000000000 0.001821494 0.000000000 0.000000000
## [2,] 0.002923977 0.002923977 0.01169591 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual      [,409]      [,410]      [,411]      [,412]      [,413]      [,414]
## [1,] 0.000000000 0.000000000 0.001821494 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.002923977 0.014619883 0.01754386 0.002923977 0.002923977
## Predicted (cv)
## Actual      [,415]      [,416]      [,417]      [,418]      [,419]      [,420]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.005847953 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual      [,421]      [,422]      [,423]      [,424]      [,425]      [,426]
## [1,] 0.000000000 0.001821494 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.000000000 0.002923977 0.002923977 0.002923977 0.00877193
## Predicted (cv)
## Actual      [,427]      [,428]      [,429]      [,430]      [,431]      [,432]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000

```

```

## [2,] 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,433] [ ,434] [ ,435] [ ,436] [ ,437] [ ,438]
## [1,] 0.001821494 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.000000000 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,439] [ ,440] [ ,441] [ ,442] [ ,443] [ ,444]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.001821494 0.001821494
## [2,] 0.002923977 0.005847953 0.002923977 0.002923977 0.000000000 0.000000000
## Predicted (cv)
## Actual [ ,445] [ ,446] [ ,447] [ ,448] [ ,449] [ ,450]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.001821494 0.000000000
## [2,] 0.002923977 0.002923977 0.002923977 0.002923977 0.000000000 0.002923977
## Predicted (cv)
## Actual [ ,451] [ ,452] [ ,453] [ ,454] [ ,455] [ ,456]
## [1,] 0.000000000 0.000000000 0.000000000 0.001821494 0.000000000 0.000000000
## [2,] 0.002923977 0.002923977 0.005847953 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,457] [ ,458] [ ,459] [ ,460] [ ,461] [ ,462]
## [1,] 0.000000000 0.001821494 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.000000000 0.002923977 0.002923977 0.002923977 0.005847953
## Predicted (cv)
## Actual [ ,463] [ ,464] [ ,465] [ ,466] [ ,467] [ ,468]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.001821494
## [2,] 0.002923977 0.002923977 0.005847953 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,469] [ ,470] [ ,471] [ ,472] [ ,473] [ ,474]
## [1,] 0.001821494 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.000000000 0.002923977 0.002923977 0.002923977 0.005847953 0.002923977
## Predicted (cv)
## Actual [ ,475] [ ,476] [ ,477] [ ,478] [ ,479] [ ,480]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,481] [ ,482] [ ,483] [ ,484] [ ,485] [ ,486]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,487] [ ,488] [ ,489] [ ,490] [ ,491] [ ,492]
## [1,] 0.000000000 0.000000000 0.000000000 0.001821494 0.000000000 0.000000000
## [2,] 0.002923977 0.005847953 0.005847953 0.000000000 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,493] [ ,494] [ ,495] [ ,496] [ ,497] [ ,498]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,499] [ ,500] [ ,501] [ ,502] [ ,503] [ ,504]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.005847953 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,505] [ ,506] [ ,507] [ ,508] [ ,509] [ ,510]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,511] [ ,512] [ ,513] [ ,514] [ ,515] [ ,516]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [ ,517] [ ,518] [ ,519] [ ,520] [ ,521] [ ,522]

```

```

## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [523] [524] [525] [526] [527] [528]
## [1,] 0.000000000 0.001821494 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.000000000 0.005847953 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [529] [530] [531] [532] [533] [534]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.005847953 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [535] [536] [537] [538] [539] [540]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.005847953 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [541] [542] [543] [544] [545] [546]
## [1,] 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000000
## [2,] 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977 0.002923977
## Predicted (cv)
## Actual [547]
## [1,] 0.000000000
## [2,] 0.002923977
##
## $prior
## [1] 0.6161616 0.3838384

## [1] "Odds ratio calculadas a partir de los coeficientes"

## (Intercept) FareIndividualBySexScaled IsMotherDummy
## 2.61338795 3.74336540 2.00399717
## LoneWolfsDummy TicketFreqScaled ageGroupsByCutAG24
## 0.76834242 0.63166906 0.21639608
## ageGroupsByCutAG36 ageGroupsByCutAG55 ageGroupsByCutAG70
## 0.14337267 0.09388892 0.02265068
## EmbarkedC EmbarkedQ PclassScaled
## 1.32332882 1.57412683 0.54614746

## [1] "Intervalos de confianza a partir de los coeficientes"

##
## Attaching package: 'questionr'

## The following object is masked from 'package:psych':
##
## describe

## [1] "Odds ratio calculadas a partir de una funcion para comprobación"

##
## Hosmer and Lemeshow goodness of fit (GOF) test
##
## data: mymod3$y, fitted(mymod3)
## X-squared = 15.684, df = 8, p-value = 0.04713

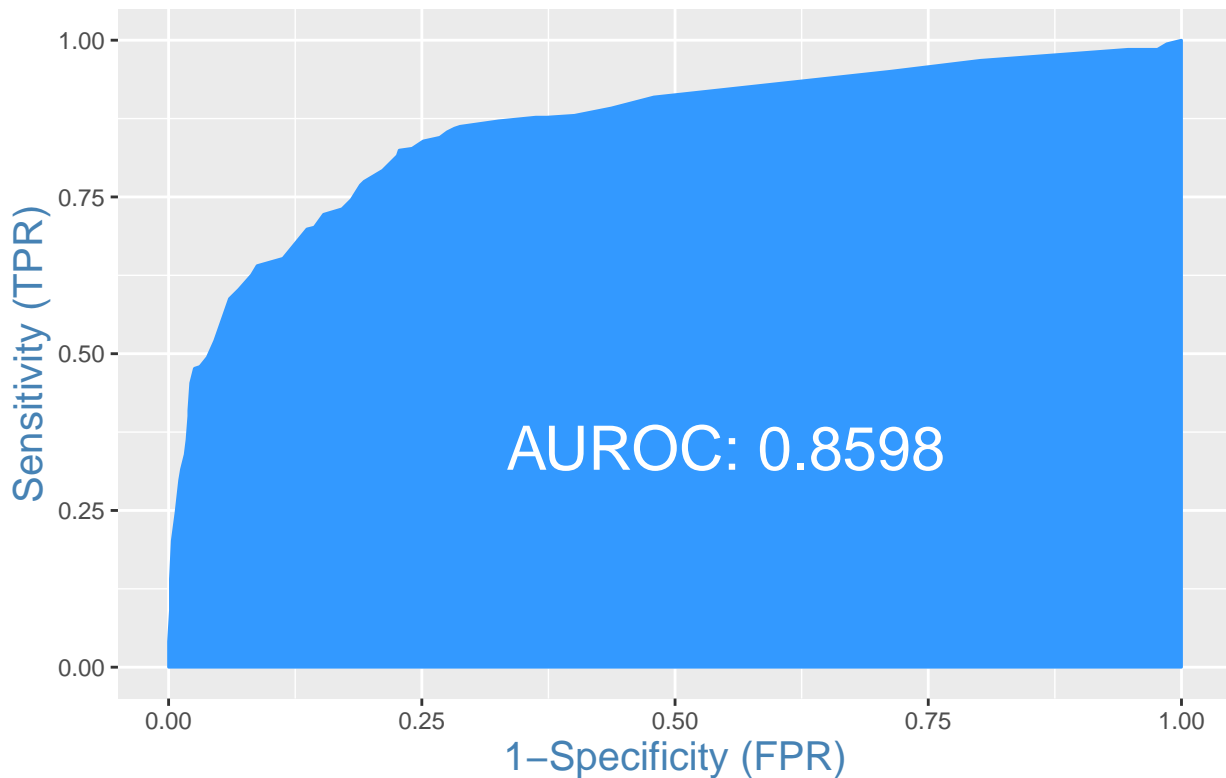
## y0 y1 yhat0 yhat1
## [0.00691,0.0705] 97 9 99.89115 6.108845
## (0.0705,0.101] 66 8 67.43804 6.561959
## (0.101,0.105] 79 9 78.97161 9.028389
## (0.105,0.151] 79 12 79.98772 11.012283
## (0.151,0.268] 77 10 69.46005 17.539949
## (0.268,0.453] 52 37 55.97799 33.022015
## (0.453,0.58] 51 38 41.97969 47.020307

```



```
## (0.58,0.717]    34 55 31.39176 57.608243
## (0.717,0.88]   10 79 17.44657 71.553430
## (0.88,0.997]   4 85  6.45542 82.544580
```

## ROC Curve



## Prediccion por redes neuronales y análisis de componetes principales

```
## Partimos el dataset en los conjuntos iniciales de train y test.
# trainCARET <- muestra[1:891,-c(4,9,11,13,14,20,54)]
# testCARET <- muestra[892:1309,-c(4,9,11,13,14,20,54)]
#
## the function createDataPartition can be used to create a stratified random sample of the data into training and testing sets.
## library(caret)
## set.seed(998)
## inTraining <- createDataPartition(trainCARET$Survived, p = .75, list = FALSE)
## trainingCARET <- trainCARET[ inTraining,]
## testingCARET <- trainCARET[-inTraining,]
#
# trainingMod<-trainCARET
# trainingMod$Survived<-as.factor(trainingMod$Survived)
# levels(trainingMod$Survived) <- c('No','Si')
# levels(trainingMod$PclassFactor) <- c('Primera','Segunda','Tercera')
# trainingMod$LoneWolfs<-as.logical(trainingMod$LoneWolfs)
# trainingMod$LoneWolfs <- ifelse(trainingMod$LoneWolfs, 'Si', 'No')
# trainingMod$IsMother<-as.logical(trainingMod$IsMother)
# trainingMod$IsMother <- ifelse(trainingMod$IsMother, 'Si', 'No')
# trainingMod$youngTravelAl<-as.logical(trainingMod$youngTravelAl)
# trainingMod$youngTravelAl <- ifelse(trainingMod$youngTravelAl, 'Si', 'No')
# trainingMod$LoneWolfs <- as.factor(trainingMod$LoneWolfs)
```

```

# trainingMod$IsMother <- as.factor(trainingMod$IsMother)
# trainingMod$youngTravelAl <- as.factor(trainingMod$youngTravelAl)
# trainingMod$Survived <- as.factor(trainingMod$Survived)
#
# testMod<-testCARET
# testMod$LoneWolfs<-as.logical(testMod$LoneWolfs)
# testMod$LoneWolfs <- ifelse(testMod$LoneWolfs, 'Si', 'No')
# testMod$IsMother<-as.logical(testMod$IsMother)
# testMod$IsMother <- ifelse(testMod$IsMother, 'Si', 'No')
# testMod$youngTravelAl<-as.logical(testMod$youngTravelAl)
# testMod$youngTravelAl <- ifelse(testMod$youngTravelAl, 'Si', 'No')
# testMod$LoneWolfs <- as.factor(testMod$LoneWolfs)
# testMod$IsMother <- as.factor(testMod$IsMother)
# testMod$youngTravelAl <- as.factor(testMod$youngTravelAl)
# testMod$Survived <- as.factor(testMod$Survived)
#
# creamos un conjunto de entrenamiento y test propios a partir de train para calcular nuestros porcentaje de acier
# ## 85% of the sample size
# smp_size <- floor(0.85 * nrow(trainingMod))
# ## Establecemos una semilla para poder volver a reproducir el mismo muestreo
# set.seed(123)
# train_ind <- sample(seq_len(nrow(trainingMod)), size = smp_size)
# training <- trainingMod[train_ind, ]
# testing <- trainingMod[-train_ind, ]
# test3Survival<-testing$Survived
# testing$Survived <- NA
#
# train_control <- trainControl(method = 'cv', number = 10, classProbs = TRUE, summaryFunction = twoClassSummary)
# #
# # # Basic Linear Modelling
# #
# # logRegModel <- caret::train(Survived ~ ., data = training, trControl = train_control,
# #                             method = 'glm', family = binomial(),
# #                             metric = 'ROC')
# # # cross validation results
# # print(logRegModel)
# # coef(summary(logRegModel))
# #
# #
# # # Naive Bayes
# # nbModel <- caret::train(Survived ~ ., data = training, trControl = train_control,
# #                          method = 'nb', tuneLength = 10,
# #                          metric = 'ROC')
# #
# # # Random Forest - Variable Importance
# # rfModel <- caret::train(Survived ~ ., data = training, trControl = train_control,
# #                         method = 'rf', tuneLength = 10,
# #                         metric = 'ROC')
# # varImp(rfModel)
# #
# # Neural Networks
# nnGrid <- expand.grid(.size = c(1,2,3,4,5,6,7),
#                      .decay = c(0, .01, .1, .2, .3, .4, .5, 1, 2))
#
# nnModel <- caret::train(Survived ~ FareIndividualBySex+IsMother+LoneWolfs+TicketFreq+ageGroupsByCut+Embarked+Pclass,
#                         method = 'nnet', tuneGrid = nnGrid,
#                         metric = 'ROC', trace = FALSE)
#

```

```

# pcaNNModel <- caret::train(Survived ~ ., data = training, trControl = train_control,
#                             method = 'pcaNNet', tuneGrid = nnGrid,
#                             metric = 'ROC', trace = FALSE)
#
# #models <- list(nn = nnModel, pcann = pcaNNModel)
# # models_preds <- lapply(models, predict, newdata = testing, type = 'prob')
# # models_probs <- as.data.frame(sapply(models_preds, function(df){1 - df$N}))
# # models_probs$Survived <- training$Survived
# nnModelTest <- data.frame(PassengerId = testing$PassengerId,
#                             Survived = ifelse(predict(nnModel, testing) == 'Si',1,0))
# pcaNNModelTest <- data.frame(PassengerId = testing$PassengerId,
#                             Survived = ifelse(predict(pcaNNModel, testing) == 'Si',1,0))
#
#
# #Vamos a comparar el resultado con las observaciones de la variable y construir una matriz de acierto
# t_cont_prediccion_contra_observación<- xtabs(~test3Survival+nnModelTest$Survived)
# print("Tabla de contingencia de predicciones survival contra observaciones survival")
# t_cont_prediccion_contra_observación
# sumatorios_filas <- rowSums(t_cont_prediccion_contra_observación)
# sumatorios_columnas <- colSums(t_cont_prediccion_contra_observación)
# total_elementos <- sum(t_cont_prediccion_contra_observación)
# # obtenemos los porcentajes de aciertos de 1s(Sensitivity) y 0s(specificity)
# porcentaje_de_acierto <- t_cont_prediccion_contra_observación[,1]/sumatorios_filas
# porcentaje_de_acierto[2]<-1-porcentaje_de_acierto[2]
# porcentaje_de_acierto
# print("Matriz de porcentaje de aciertos del modelo")
# matriz_aciertos<-cbind(t_cont_prediccion_contra_observación,porcentaje_de_acierto)
# print("acierto general del modelo")
# acierto_tortal <- (t_cont_prediccion_contra_observación[1,1]+t_cont_prediccion_contra_observación[2,2])/total_el
# acierto_tortal
# kable(matriz_aciertos, caption = paste("Matriz de acierto del modelo. acierto global:",acierto_tortal))%>%
# kable_styling(latex_options = c("striped", "hold_position"), position = "center", font_size = 8)%>%
# add_header_above(c(" ", "(Prediccion survival)" = 2," "), bold = T, italic = T)%>%
# pack_rows("(Observacioens survival)", 1, 2)
#
# #Vamos a comparar el resultado con las observaciones de la variable y construir una matriz de acierto
# t_cont_prediccion_contra_observación<- xtabs(~test3Survival+pcaNNModelTest$Survived)
# print("Tabla de contingencia de predicciones survival contra observaciones survival")
# t_cont_prediccion_contra_observación
# sumatorios_filas <- rowSums(t_cont_prediccion_contra_observación)
# sumatorios_columnas <- colSums(t_cont_prediccion_contra_observación)
# total_elementos <- sum(t_cont_prediccion_contra_observación)
# # obtenemos los porcentajes de aciertos de 1s(Sensitivity) y 0s(specificity)
# porcentaje_de_acierto <- t_cont_prediccion_contra_observación[,1]/sumatorios_filas
# porcentaje_de_acierto[2]<-1-porcentaje_de_acierto[2]
# porcentaje_de_acierto
# print("Matriz de porcentaje de aciertos del modelo")
# matriz_aciertos<-cbind(t_cont_prediccion_contra_observación,porcentaje_de_acierto)
# print("acierto general del modelo")
# acierto_tortal <- (t_cont_prediccion_contra_observación[1,1]+t_cont_prediccion_contra_observación[2,2])/total_el
# acierto_tortal
# kable(matriz_aciertos, caption = paste("Matriz de acierto del modelo. acierto global:",acierto_tortal))%>%
# kable_styling(latex_options = c("striped", "hold_position"), position = "center", font_size = 8)%>%
# add_header_above(c(" ", "(Prediccion survival)" = 2," "), bold = T, italic = T)%>%
# pack_rows("(Observacioens survival)", 1, 2)

## para envio

```

```

# Partimos el dataset en los conjuntos iniciales de train y test.
trainCARET <- muestra[1:891,-c(4,9,11,13,14,20,54)]
testCARET <- muestra[892:1309,-c(4,9,11,13,14,20,54)]

trainingMod<-trainCARET
trainingMod$Survived<-as.factor(trainingMod$Survived)
levels(trainingMod$Survived) <- c('No','Si')
levels(trainingMod$PclassFactor) <- c('Primera','Segunda','Tercera')
trainingMod$LoneWolfs<-as.logical(trainingMod$LoneWolfs)
trainingMod$LoneWolfs <- ifelse(trainingMod$LoneWolfs, 'Si', 'No')
trainingMod$IsMother<-as.logical(trainingMod$IsMother)
trainingMod$IsMother <- ifelse(trainingMod$IsMother, 'Si', 'No')
trainingMod$youngTravelAl<-as.logical(trainingMod$youngTravelAl)
trainingMod$youngTravelAl <- ifelse(trainingMod$youngTravelAl, 'Si', 'No')
trainingMod$LoneWolfs <- as.factor(trainingMod$LoneWolfs)
trainingMod$IsMother <- as.factor(trainingMod$IsMother)
trainingMod$youngTravelAl <- as.factor(trainingMod$youngTravelAl)
trainingMod$Survived <- as.factor(trainingMod$Survived)

testMod<-testCARET
testMod$LoneWolfs<-as.logical(testMod$LoneWolfs)
testMod$LoneWolfs <- ifelse(testMod$LoneWolfs, 'Si', 'No')
testMod$IsMother<-as.logical(testMod$IsMother)
testMod$IsMother <- ifelse(testMod$IsMother, 'Si', 'No')
testMod$youngTravelAl<-as.logical(testMod$youngTravelAl)
testMod$youngTravelAl <- ifelse(testMod$youngTravelAl, 'Si', 'No')
testMod$LoneWolfs <- as.factor(testMod$LoneWolfs)
testMod$IsMother <- as.factor(testMod$IsMother)
testMod$youngTravelAl <- as.factor(testMod$youngTravelAl)
testMod$Survived <- NULL
levels(testMod$PclassFactor) <- c('Primera','Segunda','Tercera')

training <- trainingMod
testing <- testMod

train_control <- trainControl(method = 'cv', number = 10, classProbs = TRUE, summaryFunction = twoClassSummary)

#Neural Networks
nnGrid <- expand.grid(.size = c(1,2,3,4,5,6,7),
                      .decay = c(0, .01, .1, .2, .3, .4, .5, 1, 2))

nnModel <- caret::train(Survived ~ FareIndividualBySex+IsMother+LoneWolfs+TicketFreq+ageGroupsByCut+Embarked+Pclass,
                        method = 'nnet', tuneGrid = nnGrid,
                        metric = 'ROC', trace = FALSE)

pcaNNModel <- caret::train(Survived ~ ., data = training, trControl = train_control,
                          method = 'pcaNNet', tuneGrid = nnGrid,
                          metric = 'ROC', trace = FALSE)

#models <- list(nn = nnModel, pcann = pcaNNModel)
# models_preds <- lapply(models, predict, newdata = testing, type = 'prob')
# models_probs <- as.data.frame(sapply(models_preds, function(df){1 - df$N}))
# models_probs$Survived <- training$Survived
nnModelTest <- data.frame(PassengerId = testing$PassengerId,
                          Survived = ifelse(predict(nnModel, testing) == 'Si',1,0))
pcaNNModelTest <- data.frame(PassengerId = 892:1309,
                             Survived = ifelse(predict(pcaNNModel, testing) == 'Si',1,0))

```

```
write.csv(nnModelTest, 'Pred_nnModel.csv',row.names = FALSE)
write.csv(pcaNNModelTest, 'Pred_pcaNNModel.csv',row.names = FALSE)
```

# 5. Representación de los resultados a partir de tablas y gráficas.

# 6. Resolución del problema. A partir de los resultados obtenidos, ¿cuáles son las conclusiones? ¿Los resultados permiten responder al problema?